



### Rapport de de projet

Niveau : Master Monnaie, Banque, Finance et Assurance

Option : Ingénierie Financière & Modélisation (IFIM)

---

# Estimation des déterminants des salaires à partir d'une fonction de type Mincer

---

*Réalisé par :*

M. NANGO FOFANA  
M. BELKHOUANE MOHAMED

*Encadré par :*

Mr. FRANCISCO SERRANITO  
(Professeur des Universités, USPN)

*Rendu fin Janvier 2022*

Promotion : 2021 - 2022

# Table des matières

<b>1 Présentation théorique de la méthodologie de Meurs &amp; Ponthieux et du modèle de Mincer.</b>	<b>iv</b>
1.1 Présentation théorique de la méthodologie de Meurs & Ponthieux. . . . .	iv
1.2 Présentation du modèle de Mincer . . . . .	v
<b>2 Visualisation de la base de données et des variables pertinentes expliquant les salaires des deux pays en fonction du genre.</b>	<b>vi</b>
2.1 Pré-traitement de la base de données des deux pays. . . . .	vi
2.2 Visualisation de la base données des deux pays. . . . .	vi
2.2.1 Visualisation de la répartition du salaire. . . . .	vi
2.2.2 Visualisation de la répartition des individus. . . . .	vii
2.3 Visualisation de la répartition des individus en fonction de la profession. . . . .	viii
2.4 Visualisation du nombres d'individus en fonction niveau d'étude . . . . .	viii
2.5 Répartition du type de contrat en fonction des individus. . . . .	ix
2.5.1 Visualisation de la répartition des individus en fonction de leurs situations.	x
2.6 Visualisation des variables pertinentes expliquant les salaires des deux pays. . . . .	x
2.6.1 Visualisation des individus (genre) en fonction du salaire. . . . .	x
2.6.2 Visualisation de la situation de l'individu en fonction du salaire. . . . .	xi
2.6.3 Visualisation de la taille de l'entreprise expliquant le salaire en fonction du genre. . . . .	xii
2.6.4 Visualisation du niveau d'étude en fonction le salaire du genre. . . . .	xii
2.6.5 Visualisation du type de contrat en fonction du salaire. . . . .	xiii
2.6.6 Visualisation du secteur en fonction du salaire. . . . .	xiv
2.6.7 Visualisation de la satisfaction en fonction du salaire. . . . .	xiv
2.6.8 Visualisation de l'ancienneté en fonction du salaire. . . . .	xv
2.7 Matrice de Corrélation des variables quantitatives. . . . .	xv
<b>3 Analyse des données des deux pays : Analyse factorielle des données mixtes (AFMD).</b>	<b>xvii</b>
3.1 Objectif . . . . .	xvii
3.2 Choix du nombre d'axes pour l'analyse et l'interprétation . . . . .	xvii
3.3 Visualisation et interprétation des variables . . . . .	xviii
3.4 Contribution des variables sur les différents axes . . . . .	xviii
3.5 Visualisation et interprétation des variables quantitatives . . . . .	xix
3.6 Visualisation et interprétation des variables qualitatives . . . . .	xx
3.7 Visualisation et interprétation des individus . . . . .	xxii
<b>4 Régression par les MCO des déterminants des salaires et des différences entre les 2 pays.</b>	<b>xxiv</b>
4.1 Objectif. . . . .	xxiv

## **Table des matières**

---

4.2 Première régression par la MCO de tous les individus des deux pays. . . . .	xxiv
4.3 Deuxième régression par les MCO utilisant l'équation de gain de Mincer de tous les individus. . . . .	xxvii
4.4 Les différents types de tests de validité du modèle associé à la première régression . . . . .	xxviii
4.4.1 Test de Chow . . . . .	xxviii
4.4.2 Test de normalité . . . . .	xxviii
4.4.3 Test d'homoscédasticité vs hétéroscléasticité . . . . .	xxix
<b>A</b>	<b>xxxii</b>
<b>Bibliographie</b>	<b>lxxv</b>

## INTRODUCTION GÉNÉRALE

Le nombre de personnes scolarisé n'a cessé d'augmenter au cours des dernières décennies moins de 5% des élèves d'une classe d'âge accèdent au baccalauréat en 1950 contre 79,6% en 2017. Cette nette augmentation appelée massification scolaire a pour objectif de faire de ces étudiants des personnes plus qualifiés. Les étudiants ont pour objectif d'obtenir un emploi à la fin de leurs formations la question que l'on se pose alors est qu'est ce qui va déterminer notre salaire ? Nous avons pour objectif d'analyser et d'identifier quels sont les déterminants du salaire. Pour cela, nous avons une base de données qui est un questionnaire, les questions posées sont en rapport avec leur emploie nous allons donc analyser les réponses de ce questionnaire et trouver et expliquer quelles sont les variables qui ont un impact sur notre salaire. Par ailleurs, nous mesurerons aussi l'impact de ces variables sur le salaire. Nous allons voir quelles sont les variables pertinentes qui expliquent les différences salariales entre chaque individu. Cette base de données recense des données de plusieurs individus ainsi que des individus travaillant dans le monde entier. Nous nous intéresserions à la France et à un pays avec un niveau de développement comparable à celle-ci, nous avons choisi l'Allemagne pour essayer de trouver des disparités entre ces 2 pays. C'est avec le logiciel RStudio que nous allons analyser ces données. Tout d'abord, nous allons importer le fichier EXCEL comportant la base de données, sur le logiciel RStudio. Nous allons vérifier si le fichier ewcs\_2015, s'est importé correctement. pour cela nous allons utiliser la commande `view( )`. Afin d'aboutir à l'objectif recherché nous allons organiser notre étude en quatre chapitres :

- Le premier chapitre « **Présentation théorique de la méthodologie de Meurs & Ponthieux et du modèle de Mincer** ». Nous ferons une brève présentation de la méthodologie de Meurs & Ponthieux et du modèle de Mincer qui sera utilisé pour la régression.
- Le deuxième chapitre « **Visualisation de la base de données** ». Ici l'objectif est de faire la visualisation des données et d'interpréter les graphiques afin de mieux comprendre nos variables.
- Le troisième chapitre « **Analyse factorielle des données mixte (AFMD)** ». Dans ce paragraphe l'objectif est de faire une analyse de manière précis de la base de données.
- Le quatrième chapitre « **Régression par les MCO(Méthode des moindres carrées ordinaire** ». Dans cette section nous ferons des régressions dans le but d'avoir une étude économétrique sur les déterminants des salaires. Pour cela nous ferons deux type régression par pays.

# **CHAPITRE 1**

## **PRÉSENTATION THÉORIQUE DE LA MÉTHODOLOGIE DE MEURS & PONTHIEUX ET DU MODÈLE DE MINCER.**

### **1.1 Présentation théorique de la méthodologie de Meurs & Ponthieux.**

Meurs et Ponthieux ont fait une analyse sur les différences salariales entre les hommes et les femmes. Elles ont constaté que l'écart de salaire mensuel moyen entre un homme et une femme est de 25,3% en 2002 soit une très faible diminution de l'écart par rapport à 1990. En effet, cet écart a diminué de seulement 1%. Elles ont cherché à montrer ces différences salariales et à les expliquer en trouvant les facteurs qui ont amené ces disparités salariales. Elles ont analysé ces facteurs tout en essayant de mettre en exergue les différentes inégalités à l'aide d'étude statistique et de test statistique. Elles se sont aussi servies de l'équation de gain (équation de mincer) pour montrer qu'à un niveau d'études et d'expériences égales il y a toujours des disparités liées à d'autres facteurs. Il y a 3 types de difficultés :

- Interruption dans leurs carrières.
- Les femmes sont plus amenées à avoir des contrats à temps partiel.
- Le secteur d'activité et la hiérarchie.

Elles ont fait leur analyse en se basant sur des données des enquêtes emploi de l'INSEE entre 1990 et 2002, à partir de 2003 l'enquête a changé et les variables n'avait plus le même sens donc elles ont décidé de choisir de faire leur analyse que sur la période 1990 et 2002. Elles ont aussi cherché à montrer les évolutions au cours du temps de ces différences salariales. Elles ont observé de faibles variations de ces inégalités avec le temps. Cette analyse s'est faite suivant la théorie du capital humain et la théorie de la discrimination. D'après la théorie du capital humain les différences de salaire est intrinsèquement liée à la productivité. L'écart de salaire entre hommes et femmes s'est décomposé en 2 parties une partie expliquée (horaire de travail inférieur chez les femmes et celle due aux autres caractéristiques tels que le secteur d'activité par exemple) et une partie non expliquée. En moyenne, les femmes font moins d'heures de travail que les hommes, elles sont plus à temps partiels. Si, nous corrigions cette variable de temps de travail en comparant des revenus horaires, nous expliquerons 8 points de l'inégalité salariale hommes-femmes qui est de 25,3%. L'inégalité non expliquée est que si l'on prend un homme et une femme exerçant le même métier avec les mêmes nombres d'heures, le même poste et qui ont un salaire différent. D'autre part, nous pouvons nous poser la question du choix du temps partiel est ce que les femmes sont davantage contraintes que les hommes à accepter le temps

partiel ?

L'écart salarial entre hommes et femmes est plus faible dans le secteur public que privé. Dans le secteur d'activité il y a des inégalités interprofessionnelles (ségrégation professionnelle) les secteurs plutôt féminins payent moins bien que les secteurs plutôt masculins. Les femmes occupent moins de postes à responsabilité, c'est l'inégalité interprofessionnelle. Malgré, avoir expliqué le temps de travail, le secteur de travail et la hiérarchie, il reste tout de même des disparités inexplicées. Meurs et Ponthieux dans leur analyse ont cherché à quantifier la part que représente chaque variable qui a un lien avec la différence salariale entre hommes et femmes par rapport à l'écart de revenus moyen entre un homme et une femme. Elles ont trouvé que le temps partiel explique 8 points de l'inégalité salariale et que le secteur d'activité explique 2 points de l'inégalité salariale. Nous allons utiliser la même méthodologie, c'est-à- dire d'identifier les variables qui peuvent expliquer les déterminants du salaire dans la fonction de type Mincer, les quantifier et les expliquer.

## 1.2 Présentation du modèle de Mincer

Jacob Mincer a voulu approfondir en 1974 la théorie du capital humain avec l'hypothèse qu'un employé avec un niveau d'éducation plus élevé aurait un salaire plus élevé. Il explique cela avec le fait qu'il aura une productivité plus élevée. Le modèle de Mincer est une équation qui a pour objectif de trouver le meilleur compromis entre continuer ces études et entrer dans le marché du travail. C'est une équation mathématique entre le nombre d'années d'étude et d'expérience ainsi qu'un autre facteur qui prend en compte les caractéristiques individuelles telles que le genre, le secteur d'activité ou le secteur public ou privé avec le logarithme du salaire. L'équation appelée équation de gains est la suivante :  $\ln \omega = \ln \omega_0 + \rho s + \beta_1 x_1 + \beta_2 x_2$ . Où le salaire est noté  $\omega$ , le salaire d'une personne sans année d'étude et d'expérience est noté  $\omega_0$ , le nombre d'année d'étude est noté  $s$ , le nombre d'année d'expérience est noté  $x$ , le taux de rendement du nombre d'année d'expérience est noté  $\beta_1$ , le taux de rendement du nombre d'année d'expérience noté  $\beta_2$ , c'est une fonction concave du nombre d'année d'expérience d'après la théorie du capital humain alors  $\beta_1$  sera positif et  $\beta_2$  sera négatif montrant que le salaire augmente avec le nombre d'année d'expérience mais avec un taux de rendement décroissant. On a le nombre d'années d'expériences optimal pour que le salaire soit maximisé est  $x^* = \beta_1 / \beta_2$ . Le capital humain est l'accumulation de connaissance et de compétence. Les études constituent un capital humain ce qui permet d'accroître la productivité. Les années d'expériences permettent d'enrichir son capital humain et d'amortir les coûts de la constitution de ce capital humain.

## CHAPITRE 2

### VISUALISATION DE LA BASE DE DONNÉES ET DES VARIABLES PERTINENTES EXPLIQUANT LES SALAIRES DES DEUX PAYS EN FONCTION DU GENRE.

#### 2.1 Pré-traitement de la base de données des deux pays.

Avant notre étude nous allons faire un pré-traitement de notre base de données. En effet notre base de données est un ensemble de questionnaire posé à un certains nombres d'individus(Hommes, Femmes) dans différents pays. Ces questions sont liées à la personnalité, son éducation, sa situation etc. Mais ces questionnaires correspondent aux variables d'études.

Pour notre études nous allons étudier deux pays qui sont la France & l'Allemagne avec pour variables : Le genre (Femme & Homme), l'âge, le salaire (Le revenu mensuel, le taux horaires, la taille de l'entreprise, la profession de l'individu, l'ancienneté (Nombre d'année d'expérience dans l'entreprise), le nombre d'heures de travail par semaine et par mois, le type de contrat (temps plein & temps partiel), la situation de l'individu (employée ou travailleurs indépendant), le secteur du travail (Privé, Public ou ONG & mixte), le niveaux d'étude, le niveau de satisfaction de l'individu et la profession de l'individu. Mais avant de rentrer dans le vif du sujet qui consiste à connaître l'écart du salaire entre l'homme et la femme et aussi dans le but de connaître le nombre d'individu nous avons vérifié et supprimer un certains nombre d'individus qui ont un salaire très éloigné du salaire moyen grâces à l'utilisation de la formule des déciles et le graphique de la boite à moustache la boite à moustache (Box-plot). Ces individus sont considérés comme des valeurs extrêmes qui peuvent qui peuvent biaiser ou fausser nos résultats. Pour cela nous allons gardé les individus qui ont un salaire supérieur ou égale à 2% et inférieur ou égale à 99% du salaire moyen. Et enfin de compte nous nous retrouvons avec 1264 individus au niveau de la base de données de la France et 1467 individus aux nivaux de l'Allemagne.

#### 2.2 Visualisation de la base données des deux pays.

##### 2.2.1 Visualisation de la répartition du salaire.

Ces deux graphiques nous montrent la répartition des salaires des deux pays.

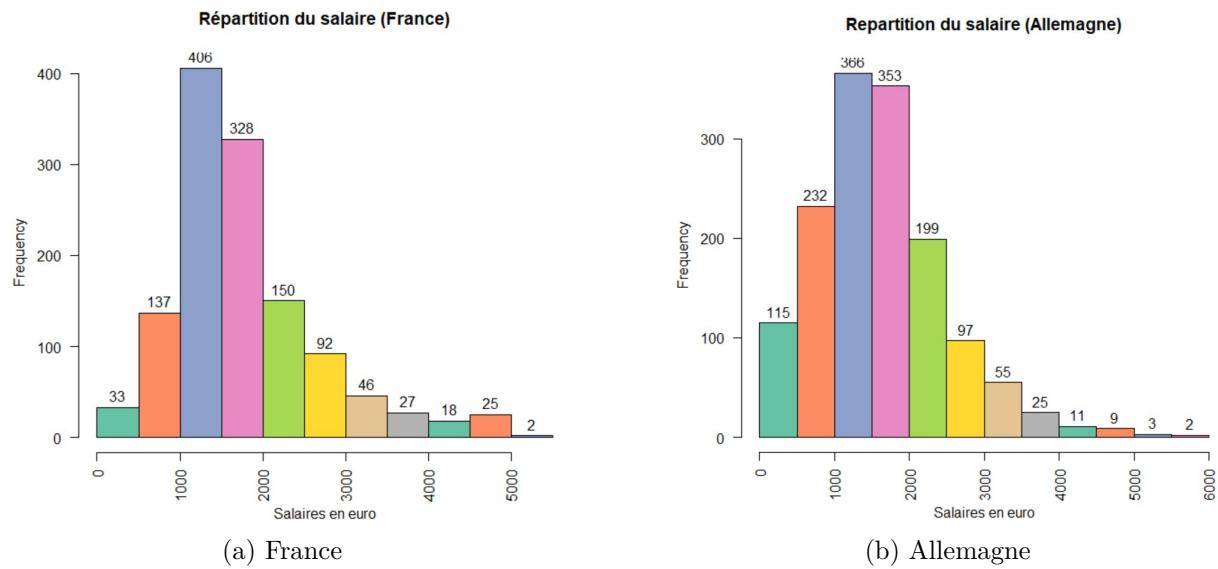


Figure 2.1 – Graphique de la répartition du salaire en fonction des individus

D'après l'histogramme des salaires nous constatons que la majorité des individus touchent en moyenne un salaire mensuel compris entre 1000 et 2000€. En effet, cette concentration du salaire entre 1000 et 2000€ s'explique par le fait qu'en France comme en Allemagne la majorité des individus touchent le SMIC qui mensuellement est de 1551,58 € en France et en Allemagne de 1584,00€. Par la suite, nous constatons que le salaire moyen en France qui vaut 1832€/mois est plus élevé que celui de l'Allemagne qui vaut 1683€/mois. De plus, 50% des individus en France touchent environ 1600 €/mois comme salaire qui est toujours plus élevé que celui de l'Allemagne qui tourne autour de 1570€/mois.

## 2.2.2 Visualisation de la répartition des individus.

Ces deux graphiques nous montre la repartitions des individus interrogés lors de notre études en fonction du genre.

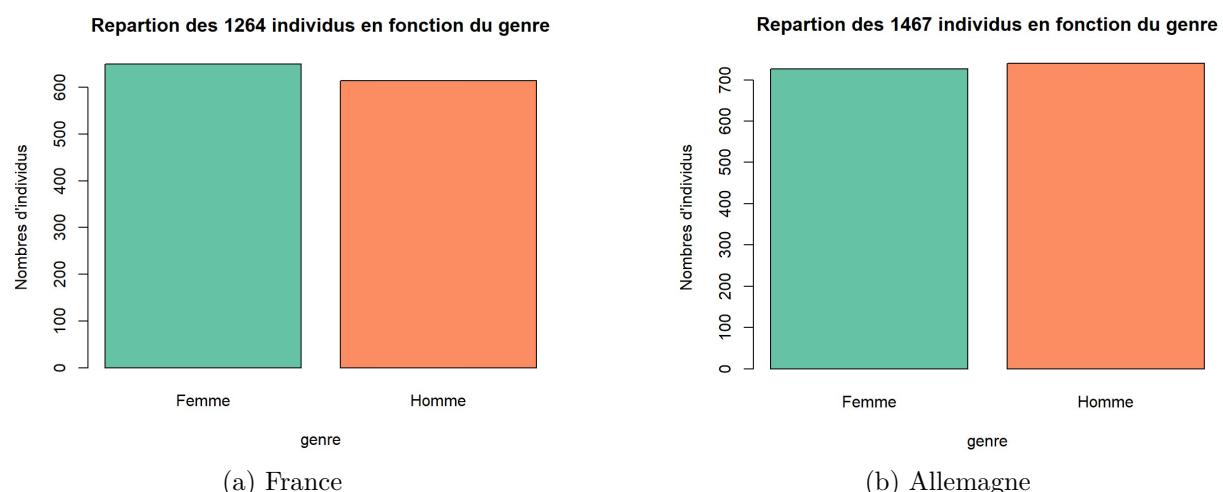


Figure 2.2 – Graphique de la répartition des individus.

D'après les graphiques nous constatons que pour notre étude nous avons plus de femmes que

## **Chapitre 2. Visualisation de la base de données et des variables pertinentes expliquant les salaires des deux pays en fonction du genre.**

d'hommes au niveau de France. D'après l'histogramme nous avons  $\approx 614$  hommes et 650 femmes soit un écart d'environ 5,86% d'homme de plus que de femme. Au niveau de l'Allemagne nous avons plus d'hommes(727) que de femmes(740). Nous avons environ  $\approx 1,80\%$  d'hommes de plus que de femmes.

### **2.3 Visualisation de la répartition des individus en fonction de la profession.**

Ces deux graphiques nous montrent la répartition des individus interroger lors de notre études en fonction de leur profession.

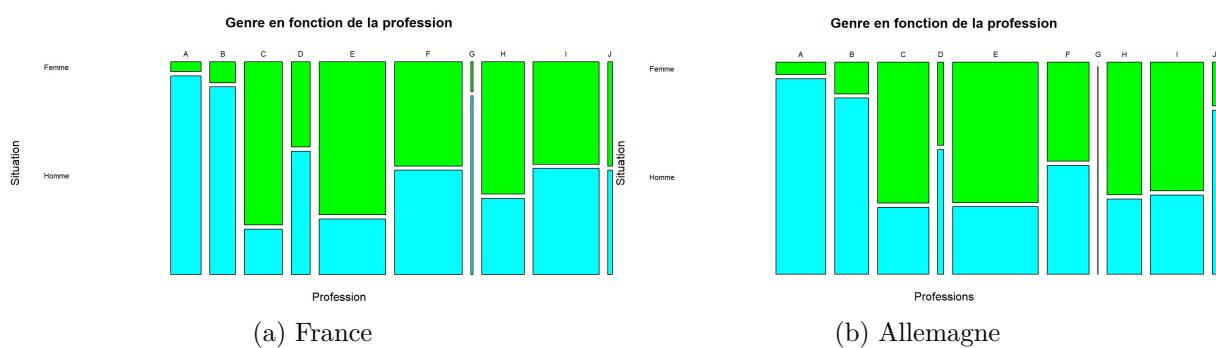


Figure 2.3 – Graphique de la répartition des profession en fonction du genre.

D'après les graphique nous constatons que sur les individus interrogés la majorité exerce le métier de Personnel de service et de vente(E), de professionnels(F) et le métier de techniciens et professionnels associés la profession(I) au niveau de la France. Parmi ces individus nous avons plus d'homme qui exercent ces professions que de femmes. Maintenant au niveau de l'Allemagne nous constatons que la majorité des individus exercent le métier de personnel de service et de vente(E) avec une majorité de féminine. Au niveau de l'Allemagne nous avons plus d'homme que de femmes.

### **2.4 Visualisation du nombres d'individus en fonction niveau d'étude .**

Dans cette section nous voulons constater la répartition des individu en fonction de leur niveau d'étude.

## Chapitre 2. Visualisation de la base de données et des variables pertinentes expliquant les salaires des deux pays en fonction du genre.

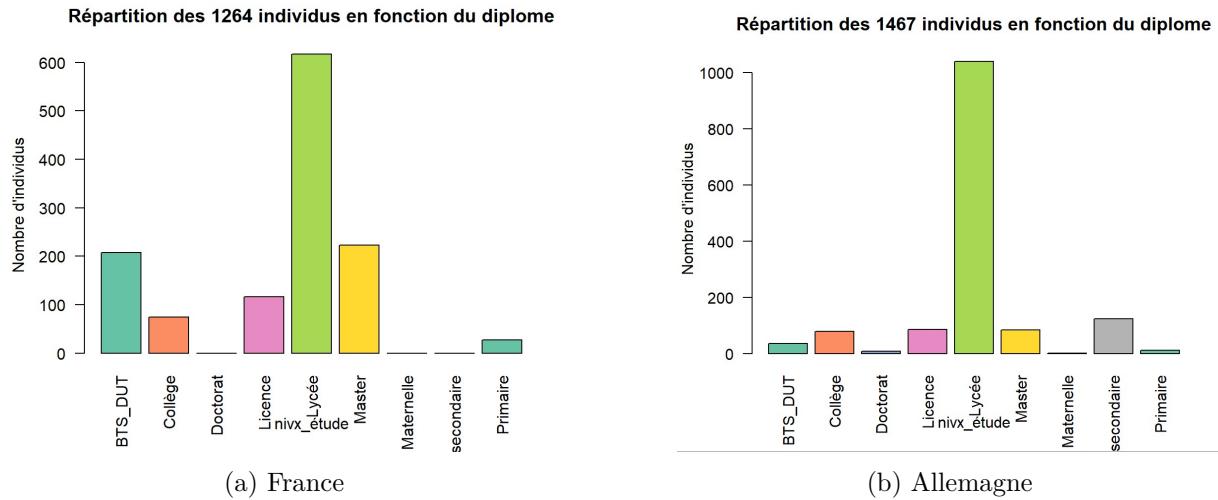


Figure 2.4 – Graphique de la répartition du niveau d'étude en fonction des individus.

On constate qu'en France comme en Allemagne, sur le nombre d'individus interrogés la majorité travail avec un niveau lycée. Parmi ces individus, nous avons environ 46.92% de femmes sur 100% et 50.81% d'hommes sur 100% en France. Au niveau de l'Allemagne, environ 70.56% de femmes sur 100% et 71.22% d'hommes sur 100% un pourcentage qui est plus élevé que celui de la France voir Annexe 1. Ensuite, contrairement à la France sur les individus interrogés, l'Allemagne possède des individus qui travaillent avec un niveau post secondaire, Maternelle et doctorat.

## 2.5 Répartition du type de contrat en fonction des individus.

Dans cette partie nous allons visualiser le nombre d'individus en fonction du type de contrat (temps plein et temps partiel) et les graphiques ci-dessous nous donne ces résultats.

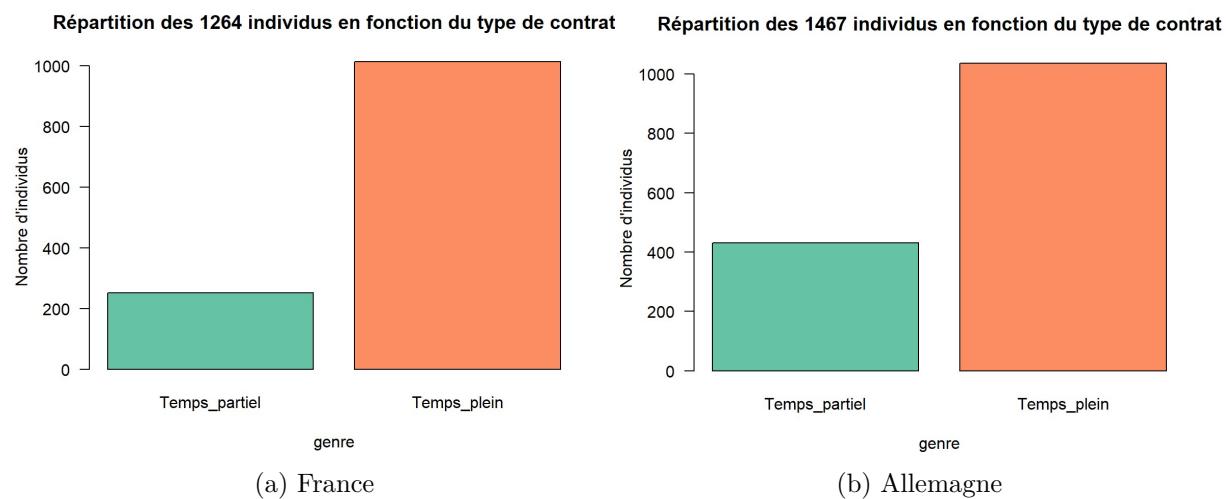


Figure 2.5 – Graphique de la répartition du type de contrat en fonction des individus.

On constate qu'en France comme en Allemagne, sur les individus interrogés durant l'étude nous avons plus d'individus qui ont un contrat de travail en temps plein soit 35 heures par semaine

que d'individu en temps partiel. Parmi ces individus interrogés, en France nous avons au niveau des femmes environ 30.31% en temps partiel et 69.69% en temps plein. Au niveau des hommes, nous avons seulement 8.79% en temps partiel et 91.21% en temps plein. Mais contrairement à la France nous avons 87.97% d'homme en temps plein en Allemagne et 12.03% en temps partiel ce nombres d'homme est plus petit que celui de la France. Quant aux femmes Allemandes interrogées 47.04% travaillent à temps partiel et 52.96% en temps complet. Pour plus de détaille voir : Annexe 1.

### **2.5.1 Visualisation de la répartition des individus en fonction de leurs situations.**

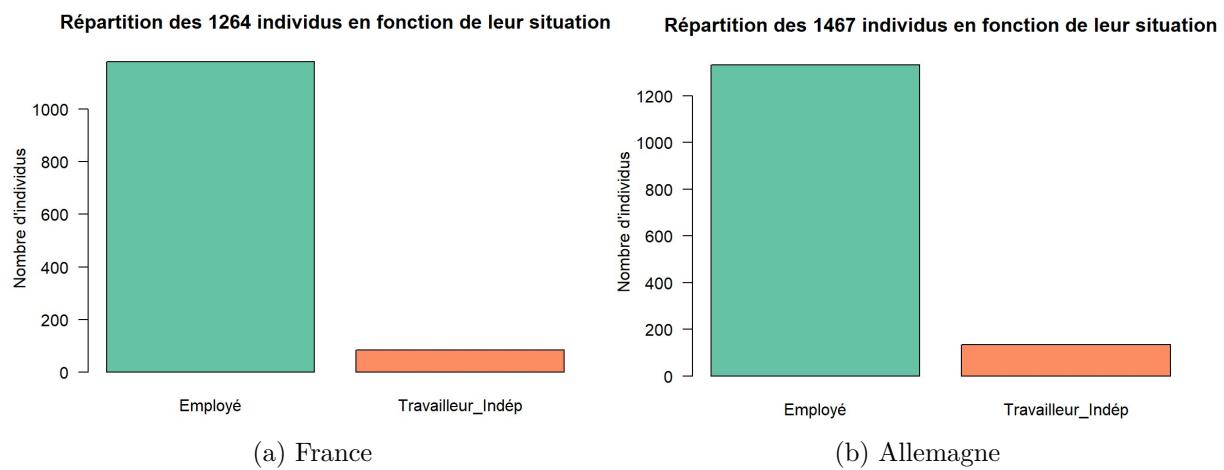


Figure 2.6 – Graphique de la répartition du de la situation de l'individu en fonction des individus.

Comme nous pouvons le constater, à partir des graphiques que ce soit l'Allemagne ou la France, la majorité des individus sont des employés. En effet, parmi ces individus sur 100% de femmes qui travaillent en France 94.92% sont des employées et 5.08% sont des travailleurs indépendants, qui travaillent dans leurs propres entreprises. Au niveau des hommes, 91.86% travaillent pour une entreprise et 8.14% travaillent pour eux. Ce qui revient à dire, que nous avons plus de femmes employées que d'hommes, il y a environ 3% de plus d'hommes qui travaillent dans leurs propres entreprises que de femmes. Quant à l'Allemagne, nous avons les mêmes constats.

## **2.6 Visualisation des variables pertinentes expliquant les salaires des deux pays.**

### **2.6.1 Visualisation des individus (genre) en fonction du salaire.**

Les graphique ci-dessous mettent en relation le salaire et le genre par rapport aux pays(France, Allemagne).

## Chapitre 2. Visualisation de la base de données et des variables pertinentes expliquant les salaires des deux pays en fonction du genre.

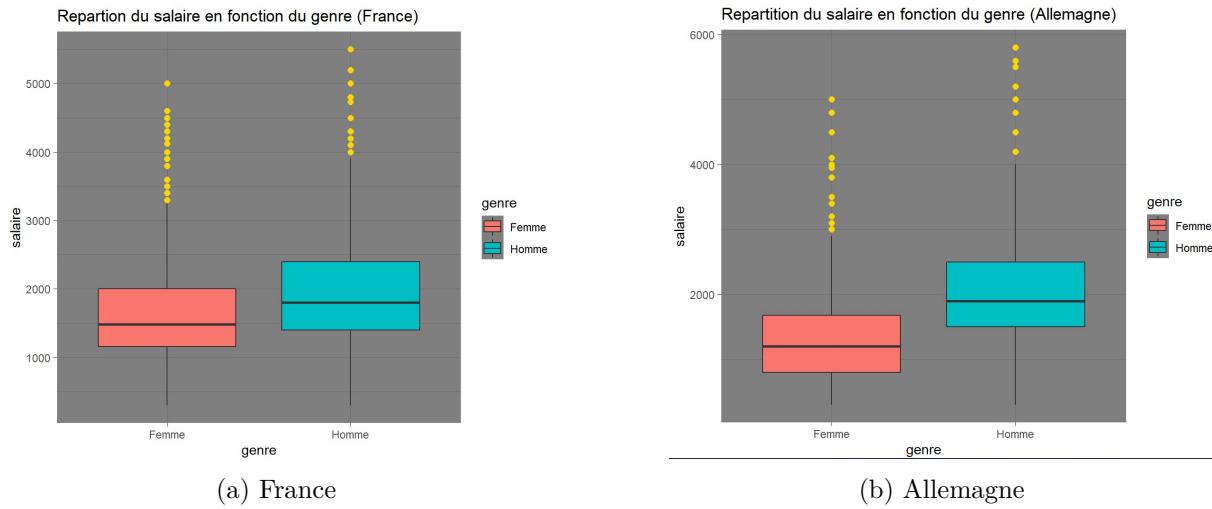


Figure 2.7 – Graphique des individus(genre) en fonction du salaire.

D'après les deux graphiques nous ne constatons que quelque soit le pays, parmi les individus interrogés les hommes ont en moyenne un salaire est plus élevés que le salaire des femmes. En France le salaire moyen des Hommes est de 2013€et celui des femmes  $\approx 1662\text{€}$ soit un écart d'environ  $\approx 21,12\%$ . Au niveau de l'Allemagne on constate un énorme écart salariale contrairement à la France d'environ  $\approx 52\%$  avec un salaire moyen des hommes d'environ 2027€et chez les femmes nous constatons un salaire moyen d'environ 1334€. Ce qui nous permet de dire que les hommes sont mieux payé que les femmes.

### 2.6.2 Visualisation de la situation de l'individu en fonction du salaire.

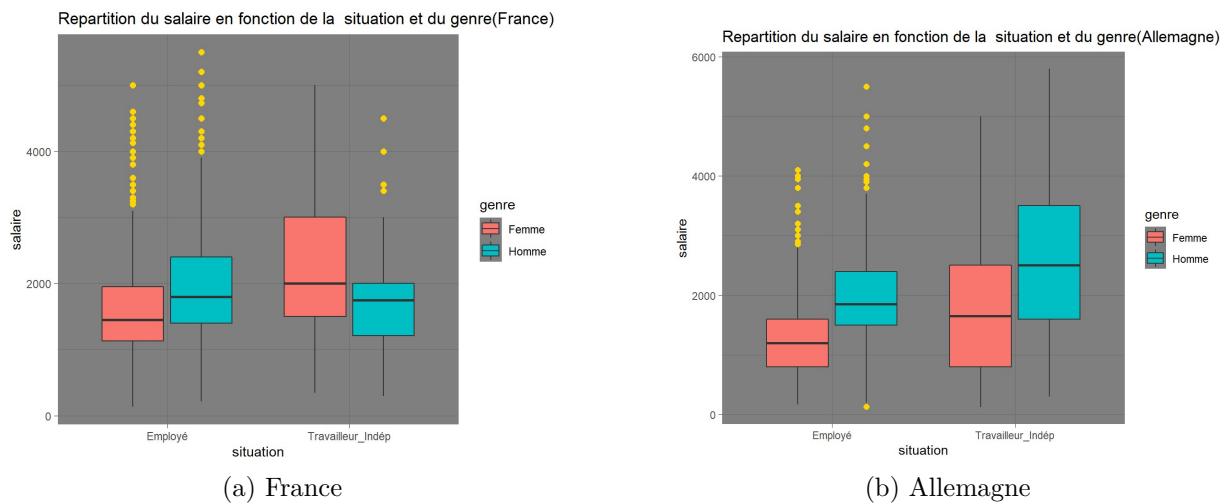


Figure 2.8 – Graphique de la situation de l'individu en fonction du salaire.

D'après les deux graphiques nous voyons que les individus qui sont des travailleurs indépendants ont en moyenne un salaire plus élevé que celui des individus employé quelque soit le pays. Au niveau de la France le salaire moyen des travailleurs indépendants s'élève à 2021€/mois et celui des employé(e)s à 1819€/mois avec pour écart de 11,10% et en Allemagne le salaire moyen des indépendants est de 2316€/mois et pour les employé(e)s de 1619€/mois soit un écart de 43,05%.

## Chapitre 2. Visualisation de la base de données et des variables pertinentes expliquant les salaires des deux pays en fonction du genre.

On remarque qu'en France parmi les individus employés, les hommes (2031 €) touchent un salaire en moyenne plus élevés que celui des femmes (1626 €) en France soit un écart de salaire d'environ  $\approx 19,94\%$  et au niveau des individus travailleurs indépendants ce sont les femmes qui touche en moyenne un salaire(2337 €) plus élevé que celui des hommes (1812 €) avec un écart salariale d'environ  $\approx 28,97\%$ . Ce qui revient à dire qu'au niveau des individus employés les hommes sont mieux payé que les hommes et au nivaux des indépendants ce sont les femmes qui sont les mieux payés. On constate qu'en Allemagne, contrairement à la France parmi les individus travailleurs indépendants ce sont les hommes qui ont un salaire (2606 €) en moyenne plus élevés que celui des femmes (1853 €) avec un écart salarial de 28,89%. Au niveau des individus employés ce sont les hommes qui ont un salaire(1954 €) plus élevés que celui des femmes(1294) avec un écart de salaire de 33,78%. De ces résultats on peut comprendre que les individus Français sont mieux payé que les Allemands.

### 2.6.3 Visualisation de la taille de l'entreprise expliquant le salaire en fonction du genre.

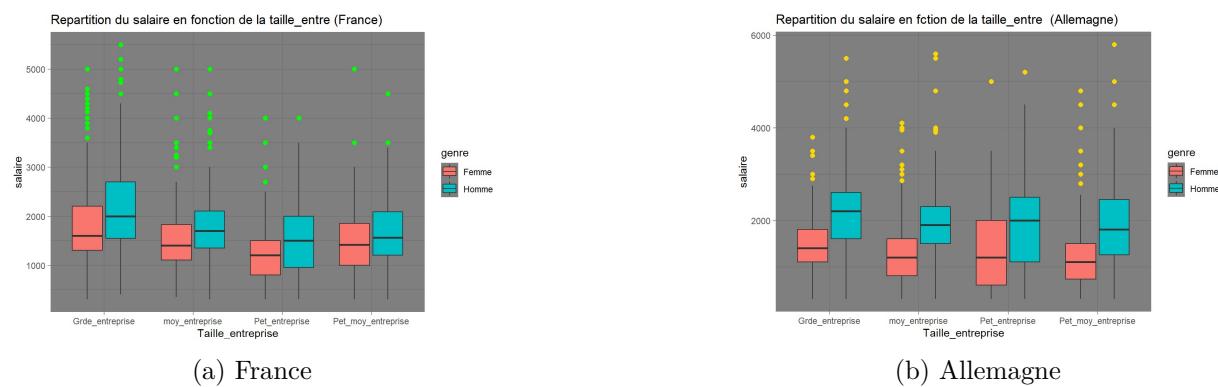


Figure 2.9 – Graphique de la taille de l'entreprise fonction du salaire.

Au niveau de la France on constate que parmi les individus travaillant dans une entreprise seul ceux qui travaillent dans une grande entreprise touchent un salaire plus élevé. Les hommes sont toujours en moyenne mieux payé (2242 €) que les femmes(1890 €). Entre ces individus nous avons un écart salarial  $\approx 15,7\%$

### 2.6.4 Visualisation du niveau d'étude en fonction le salaire du genre.

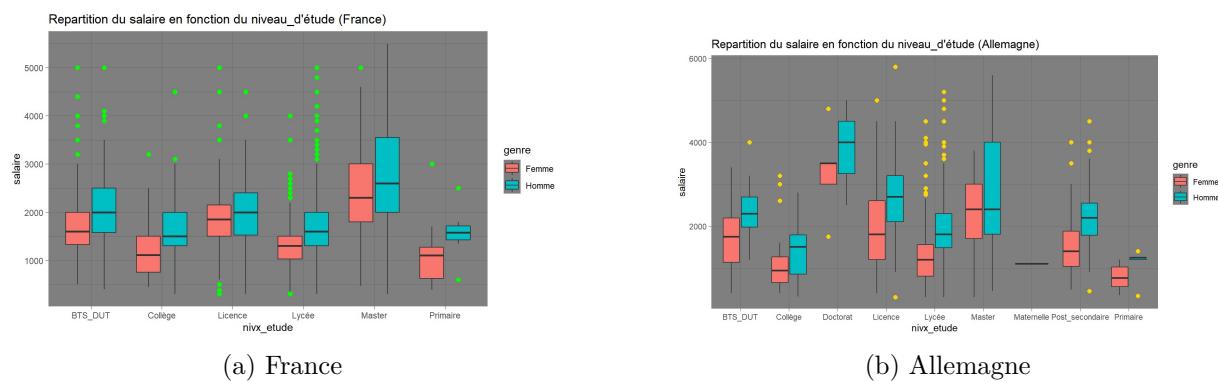


Figure 2.10 – Graphique du niveau d'étude fonction du salaire.

D'après le graphique au niveau de la France on constate que ce sont les individus ayant un niveau master qui en moyenne un salaire plus élevé que les autres niveau ce qui est normal car plus le niveau d'étude de l'individu est élevé plus le salaire est élevé. Parmi ces individus ayant un niveau master on remarque que les hommes(2832 €) touchent en moyenne un salaire plus élevé que celui des femmes(2475 €). On constate un écart salarial de 14,42%. Au niveau de l'Allemagne, les individus diplômés d'un doctorat ont un salaire en moyennes plus élevés que les autres niveaux. Parmi ces individus, les hommes sont en moyenne mieux payés(3833 €) que les femmes (3310 €) avec un écart salariale de 15,80% entre l'homme et la femme.

### 2.6.5 Visualisation du type de contrat en fonction du salaire.

Nous allons faire la visualisation de la répartition des salaires en fonction du type de contrat.

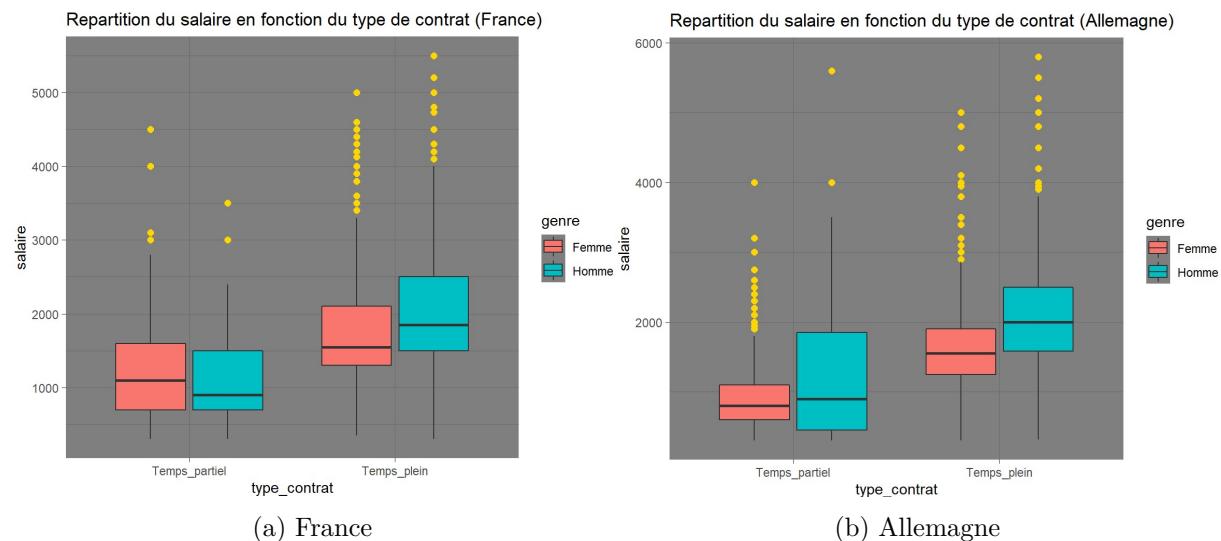


Figure 2.11 – Graphique du type de contrat en fonction du salaire

À partir des graphiques, on constate qu'en France sur l'ensemble des individus interrogés, les femmes qui travaillent en temps partiels ont en moyenne un meilleur salaire que les hommes. Mais quant aux individus qui travaillent en temps plein, on constate que ce sont les hommes qui sont le mieux payés. Les femmes qui travaillent en temps plein, sont moins bien payées que les hommes qui touchent en moyenne un salaire de 69.69% travail en temps plein. Aussi chez les hommes de femme et 8.79% des hommes travail en temps partiel qui ont en moyenne un salaire  $\approx 1600\text{€}/\text{mois}$  voir (Annexe 1 et 4). En Allemagne contrairement à la France les hommes qui travaillent en temps partiel touche un salaire plus élevé que les femmes, c'est aussi le cas des hommes qui travailles en temps plein.

### 2.6.6 Visualisation du secteur en fonction du salaire.

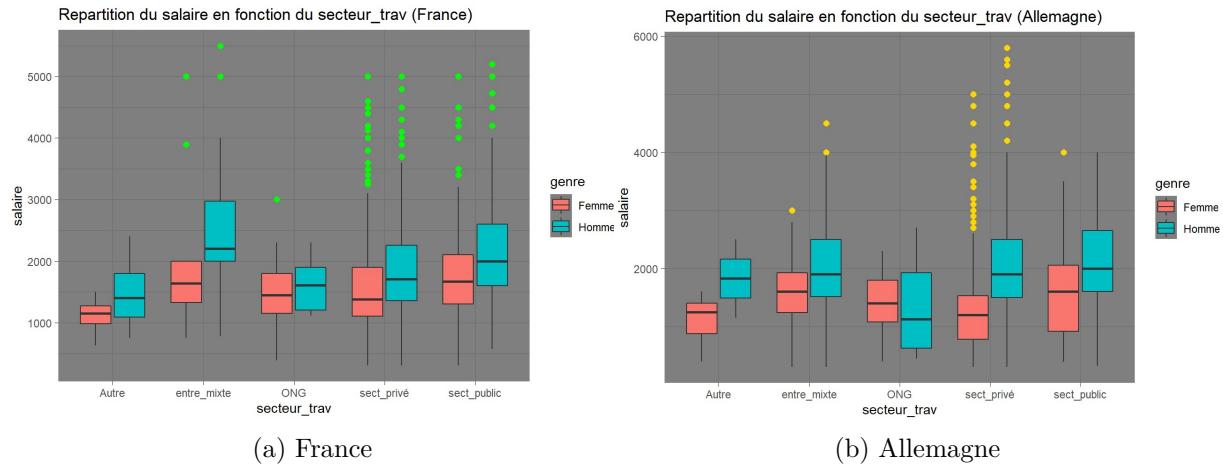


Figure 2.12 – Graphique du secteur en fonction du salaire

On constate que les individus travaillant dans une entreprise ou secteur mixte ont un salaire en moyenne plus élevés que dans les autres secteurs. On constate que dans ce secteur mixte les hommes touchent en moyenne un salaire (2671€/mois) plus élevés que le salaire des femmes (1999€/mois) avec un écart salariale d'environ 33,62%. Quant à l'Allemagne on constate que c'est le secteur mixte qui est le mieux payé et dans ce cas ce sont les hommes(2186€/mois) qui sont en moyenne un salaire majoré par rapport à celui des femmes (1585€/mois). Soit environ  $\approx 37,92\%$  d'écart salariale. Nous pouvons constater que le secteur le moins payé est le secteur "autre" dans les deux pays.

### 2.6.7 Visualisation de la satisfaction en fonction du salaire.

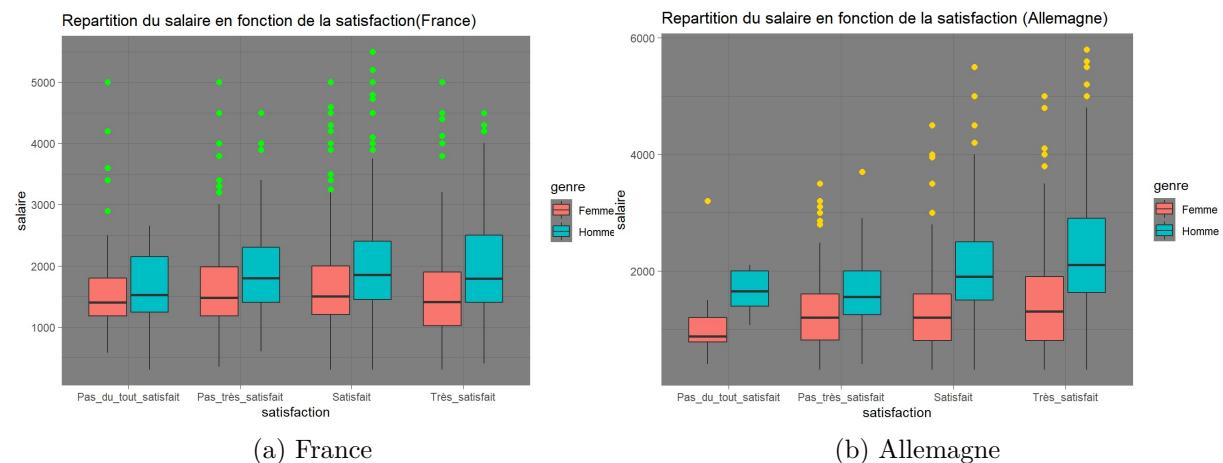


Figure 2.13 – Le salaire en fonction de la satisfaction

D'après ces graphiques nous pouvons constater que quelque soit le pays les individus sont très satisfait de leur salaire. Parmi ces individus, les hommes sont toujours plus satisfait.

### 2.6.8 Visualisation de l'ancienneté en fonction du salaire.

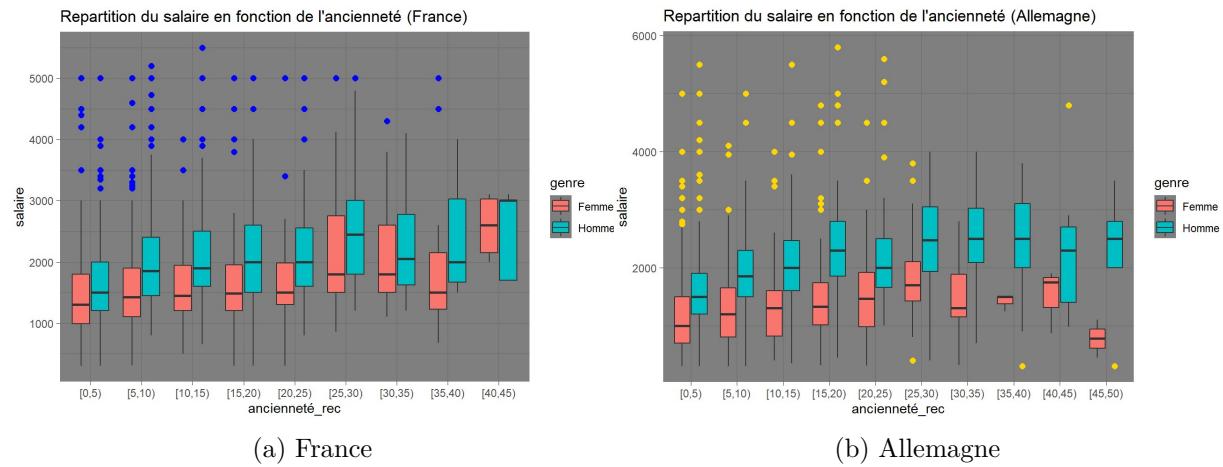


Figure 2.14 – Graphique de la satisfaction en fonction du salaire

On remarque les individus qui ont plus d'expérience ont un salaire en moyenne plus élevé que les individus qui ont moins d'expérience quelque soit le pays. Cela s'explique par le fait que plus un individu acquiert de l'expérience au fil du temps et plus il devient productif cela s'explique avec la théorie du capital humain. Mais nous constatons qu'en Allemagne les individus qui ont des années expérience comprises entre 40 ans et 50 ans sont moins bien payé ce qui n'est peut-être pas très significatif compte tenu de la taille de l'échantillon qui est assez faible. On voit aussi que dans cette situation les hommes ont un salaire toujours élevé que celui des femmes quel que soit le nombre d'année d'expérience.

### 2.7 Matrice de Corrélation des variables quantitatives.

L'objectif est de voir la corrélation entre les variables quantitatives.

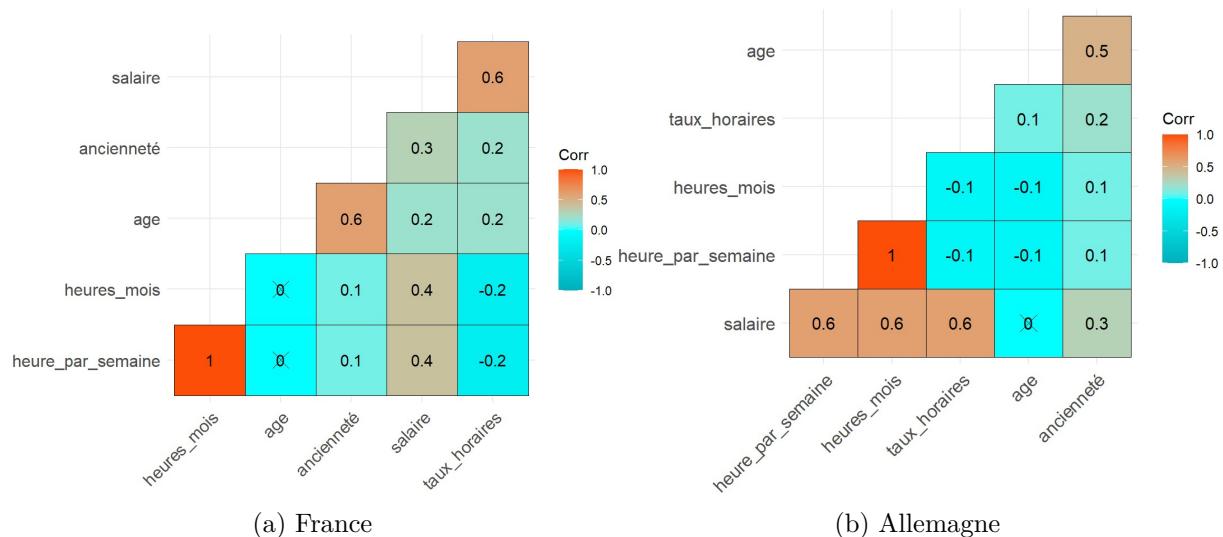


Figure 2.15 – Matrice de corrélation

**Chapitre 2. Visualisation de la base de données et des variables pertinentes expliquant les salaires des deux pays en fonction du genre.**

---

D'après la matrice de corrélation de la France on constate que le salaire est corrélé positivement soit ( $R = 0.6$ ) avec le taux horaire, on constate aussi une corrélation entre l'âge et l'ancienneté ( $R = 0.6$ ). Ceux sont uniquement ces deux variable. La corrélation de l'âge avec l'ancienneté est du au fait que plus l'âge de l'individu augmente son niveau d'expérience augmente aussi. Contrairement à la France on constate que pour l'Allemagne le salaire est corrélé positivement à la fois avec le nombre d'heure de travaille par semaine, le nombre d'heure de travaille par mois et avec le taux horaire avec un coefficient de corrélation qui vaut  $R = 0.6$ . Aussi comme dans le cas de la France l'âge et l'ancienneté sont corrélés soit environ  $R = 0.5$ .

# CHAPITRE 3

## ANALYSE DES DONNÉES DES DEUX PAYS : ANALYSE FACTORIELLE DES DONNÉES MIXTES (AFMD).

### 3.1 Objectif

L'objectif est faire une analyse multivariée de notre base de données. D'après notre recherche il existe plusieurs méthodes d'analyse de données. Pour notre analyse étant donné que notre base de données comporte des données mixtes nous utiliserons l'AFDM appelée en anglais(FAMD) qui correspond à la méthode la plus adaptée pour effectuer notre analyse. Afin d'effectuer une bonne analyse de notre base de données nous allons utiliser tous les variables choisie dans la section 2.1. Pré-traitement de la base de données des deux pays

### 3.2 Choix du nombre d'axes pour l'analyse et l'interprétation

Pour le choix du nombre d'axes pour l'analyse et l'interprétation nous allons utiliserons la méthode des coudes. Cette méthode des coudes consiste a choisir le nombre d'axes(dimension) à partir de la cassure de la représentation de la pourcentage expliqué des valeurs propres. Pour cela les graphiques ci-dessous nous permettrons de faire le choix.

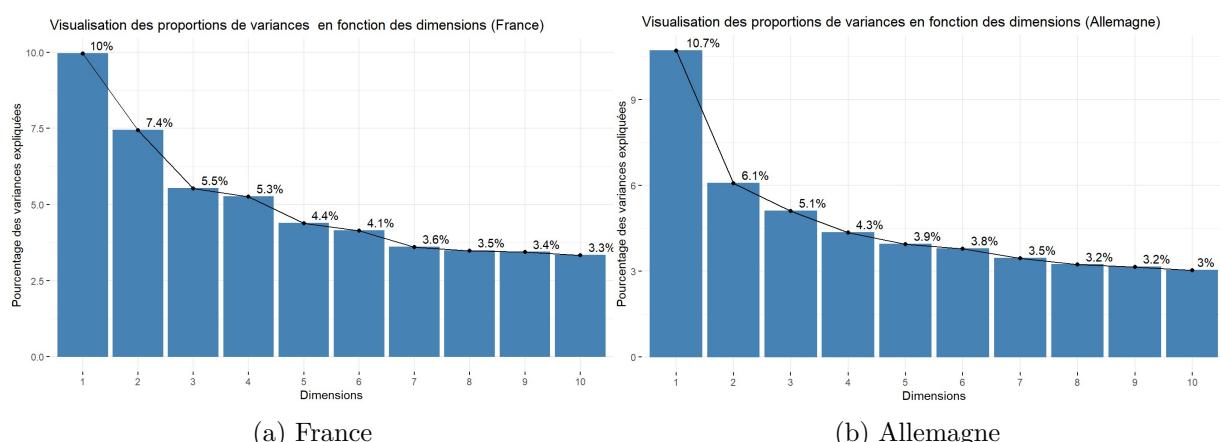


Figure 3.1 – Visualisation des valeurs propres

A partir des graphiques ci-dessus nous pouvons retenir que les deux premier axes pour notre analyses quelques soit le pays.

**Attention :** L'objectif est de choisir environ 50 % des variance expliqué pour éviter de perdre plus d'information mais, vue notre base de données nous utiliserons comme nous précisé  $\approx 20\%$  de l'inertie expliqué. Voir Annexe 8

### 3.3 Visualisation et interprétation des variables

L'objectif est de faire la représentation graphique de toutes les variables utilisé pour notre étude afin de savoir laquelle des variables expliques le mieux les axes i.e contribuent à la formation des axes.

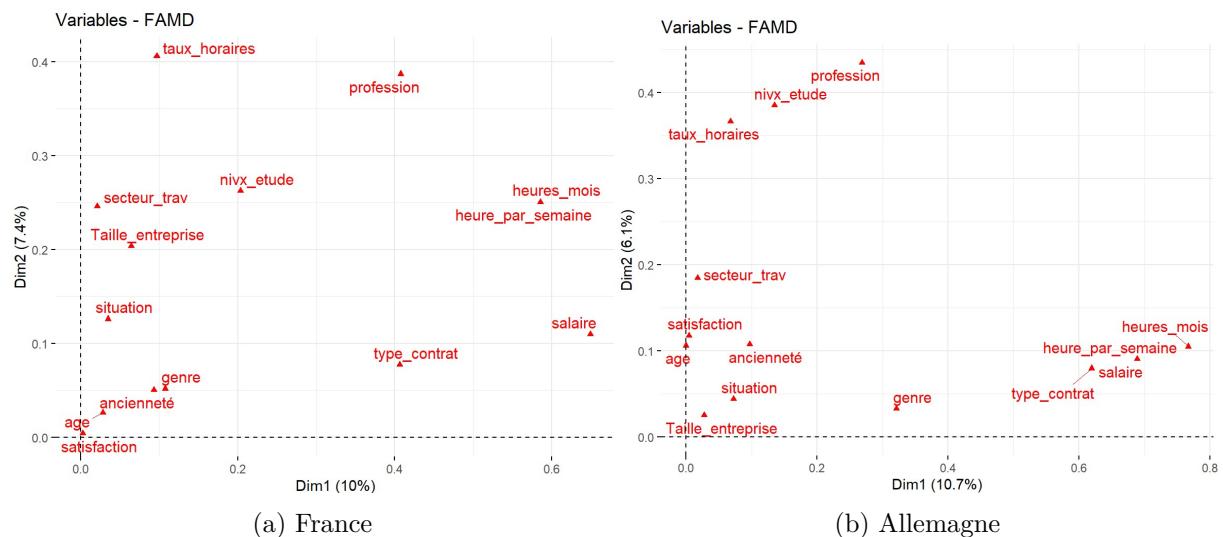


Figure 3.2 – Graphiques des variables

Au niveau de la France on constate que les variables telles que le salaire, le nombre d'heures de travail par mois, par semaine et le type de contrat sont les variables qui contribuent le plus à la formation de l'axe 1, et quant à l'axe 2 nous avons comme variables : le taux horaire, la profession, le niveaux d'étude, le secteur du travail et la taille de l'entreprise qui contribuent le plus à cet axe. Quant à l'Allemagne on remarque que les variables qui contribuent à la formation de l'axe 1 sont : le nombre d'heure de travail par mois, par semaine, le salaire, le genre et le type de contrat. Au niveau de l'axe 2, la profession le niveau d'étude d'étude, le taux horaire et le secteur du travail sont les variables qui contribuent le plus a la formation de l'axe 2.

Afin de confirmer la contribution des variables à la formation des différentes axes nous allons dans la section suivantes faire les graphique de la contribution des variables pour être rassuré.

### 3.4 Contribution des variables sur les différents axes

Dans cette section nous ferons des histogramme pour visualiser la contribution des variables.

### Chapitre 3. Analyse des données des deux pays : Analyse factorielle des données mixtes (AFMD).

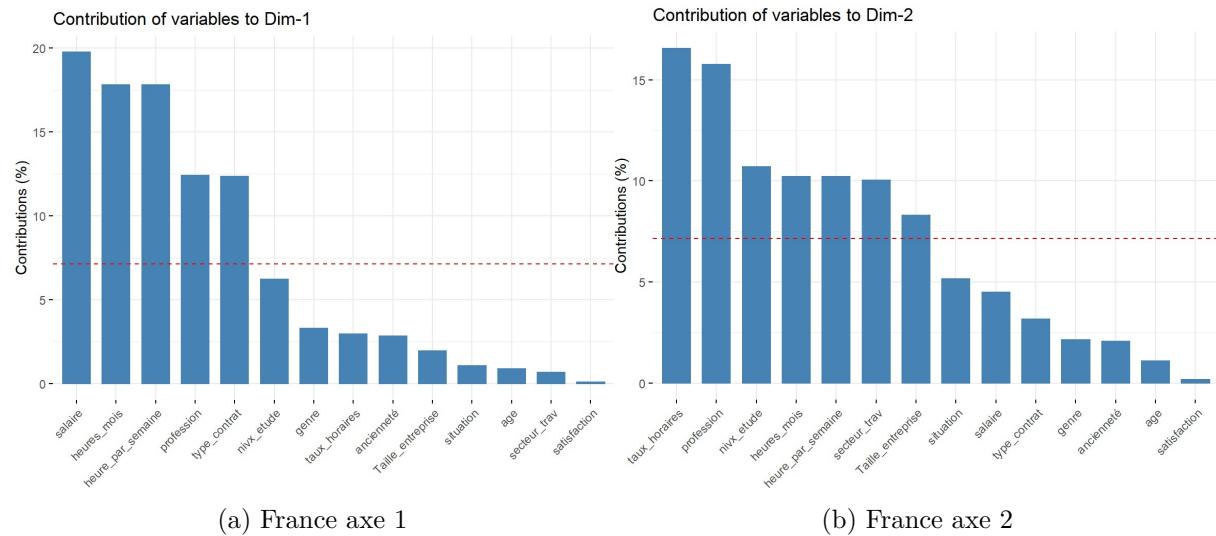


Figure 3.3 – Contribution des variables sur les différents axes(France)

A partir de ces histogramme nous pouvons confirmer notre hypothèse. En effet, on constate que les variable qui contribuent aux différents axes ont leurs valeurs au dessus de la valeur moyenne attendue(trait rouge). Nous pouvons dire que les variables qui contribuent le plus sont ces variables qui influence le salaire.

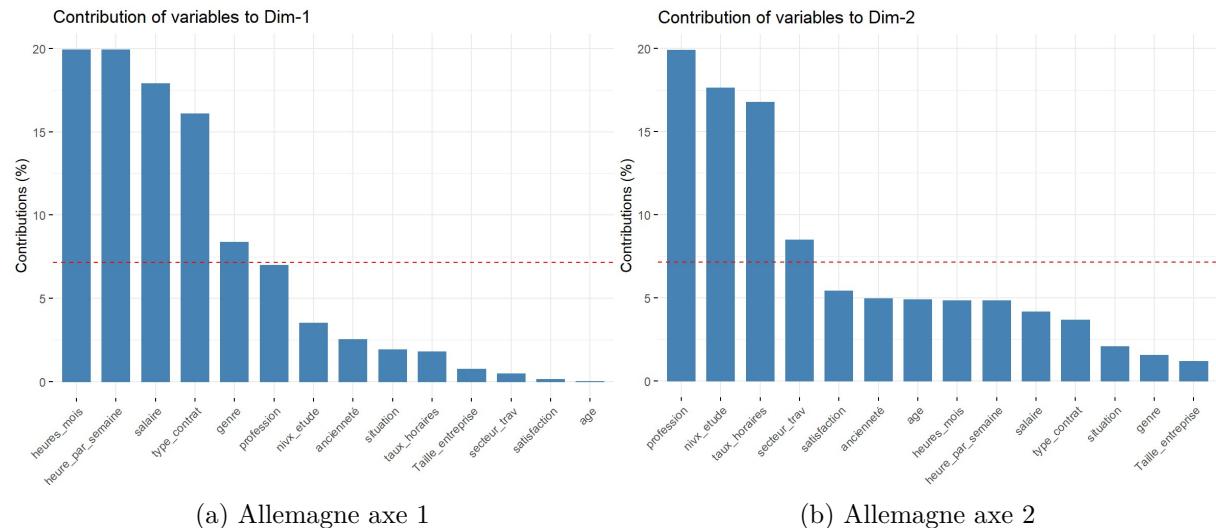


Figure 3.4 – Contribution des variables sur les différents axes(Allemagne)

Comme pour la France le graphique de contribution des variables confirme notre hypothèse. Voir Annexes 9

### 3.5 Visualisation et interprétation des variables quantitatives

Dans cette section, nous décrirons comment visualiser les variables quantitatives. Cette visualisation n'est rien d'autre que le cercle de corrélation des variables quantitatives. Ce cercle de corrélation nous permettra de connaître la qualité de représentation des variables.

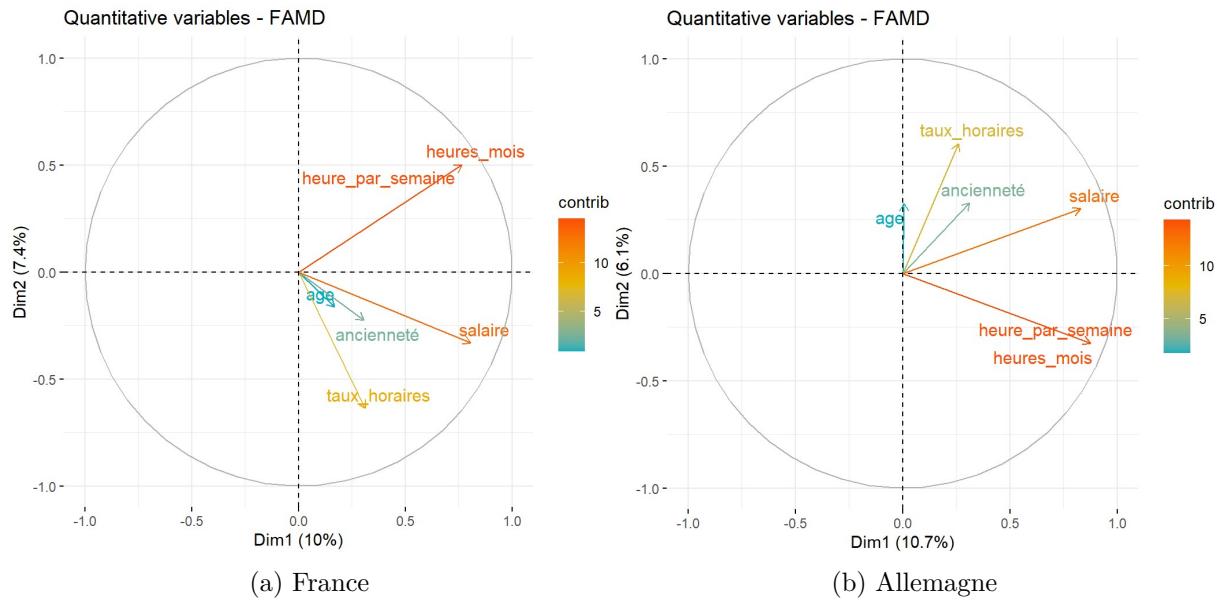


Figure 3.5 – Graphiques des variables quantitatives

La représentation du cercle des corrélations des variables montre qu'un grand nombre de variables sont plutôt bien représentées car proches du cercle. On remarque qu'en France comme en Allemagne, le salaire, le taux horaire et les nombres d'heures (mois et semaine) de travail sont bien représentés dans le graphique. Contrairement aux autres variables telle que l'âge et l'ancienneté sont en revanche beaucoup moins bien représentés, car les flèches sont courtes.

**Interprétation :** En effet, nous constatons qu'en France comme en Allemagne le salaire, le nombre d'heure par mois et par semaine sont des variables fortement corrélées positivement avec l'axe 1 (0.6 à 0.9) qui sont très importantes dans la formation de la première dimension. En effet, la corrélation forte de ses variables s'expliquent par le fait que le nombre d'heure de travail par semaine de l'individu augmente automatiquement le nombre d'heure de travail par mois car celui-ci est lié aux semaines, par conséquent le salaire mensuel augmente également. Au niveau de l'axe deux (2) le taux horaire de la France est corrélé négativement et est la seul variable au niveau de l'Allemagne corrélé positivement. De plus, le nombre d'heure par mois et par semaine sont corrélés positivement avec l'axe 2 uniquement pour la France. La corrélation négative du taux horaire avec l'axe deux s'explique par le fait que plus le nombre d'heure de travail par mois augmente moins on gagne car plus le volume horaire est élevé le taux horaire.

### 3.6 Visualisation et interprétation des variables qualitatives

Nous allons faire la représentation graphique des variables qualitatives. L'objectif de ces représentations consiste à mieux voir la contribution des variables qualitatives afin de confirmer nos hypothèses.

### Chapitre 3. Analyse des données des deux pays : Analyse factorielle des données mixtes (AFMD).

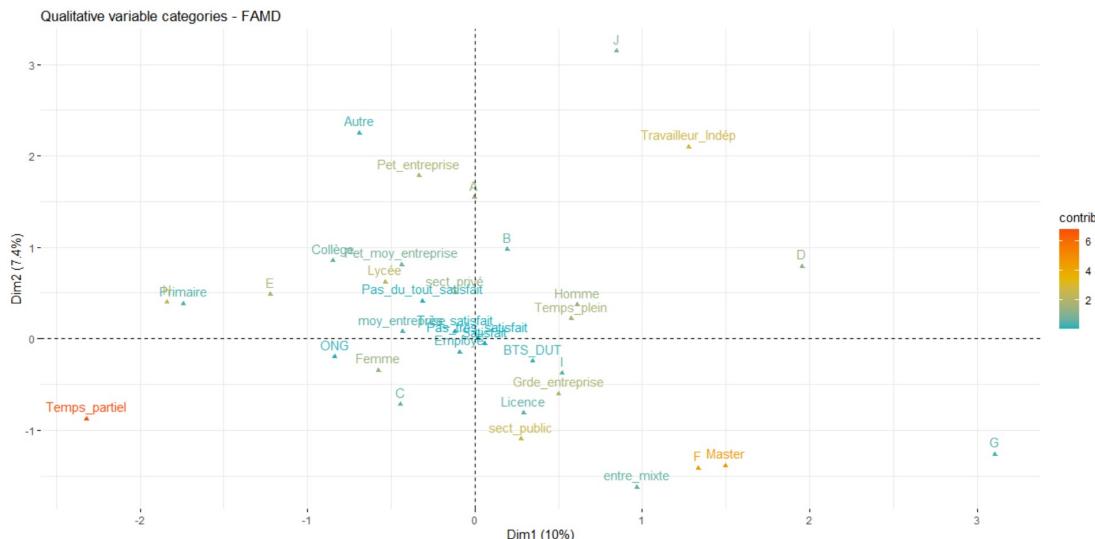


Figure 3.6 – Graphique des variables qualitatives France

D'après le graphique ont constaté que les variables qualitative qui contribuent le plus à la formation des deux axes sont la profession, le type de contrat et le niveau d'étude pour l'axe 1 et pour l'axe 2 nous avons le secteur du travail, la situation de l'individu et aussi la profession. Ces résultats confirme nos hypothèses de la section sur la représentation graphiques des résultats.

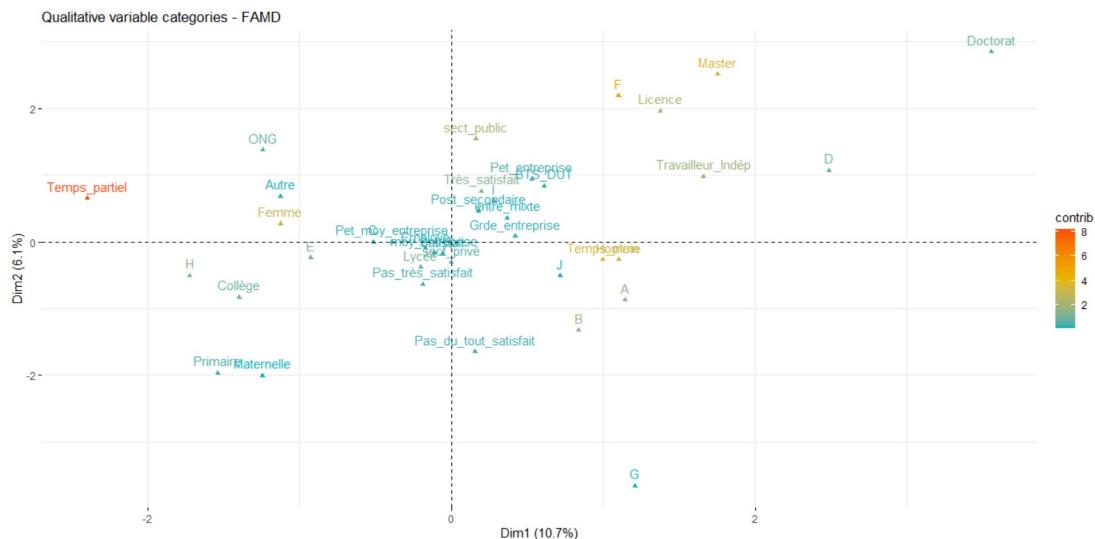


Figure 3.7 – Graphique des variables qualitatives Allemagne

Au niveau de l'Allemagne on a le type de contrat la profession et le niveau d'étude qui contribuent à la formation de l'axe 1 et pour l'axe 2 nous avons la profession le secteur du travail qui ont contribué le plus à la formation de cet axe. Nos résultats sont confirmé quant à la contribution des variables.

### 3.7 Visualisation et interprétation des individus

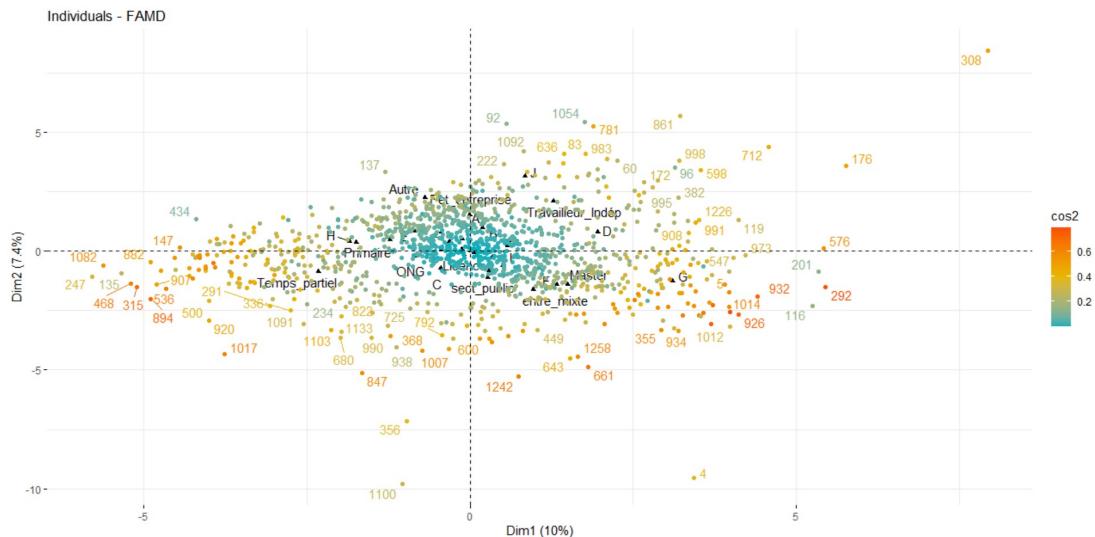


Figure 3.8 – Graphique des individus France

**Interprétation :** Ce graphique nous montre la représentation graphique des individus en couleurs et des variables qualitatives en noirs. On voit bien presque tous les individus sont proche les uns des autres proches ce qui permet de savoir si nous avons des individus qui se ressemblent. Aussi certains individus sont plus proche des variables qualitatives . On voit que les individus 336, 234, 291 sont proches de la variables temps partiel ce qui revient a dire que ces individus travaillent en temps partiel. De plus nous constatons que les individus colorés en vert sont proche des variables suivantes : profession(H,C,A,F,I, D), le niveau master et licence, primaire, secteur public ou ONG, travailleur indépendant, petite entreprise, entreprise mixte. Cette liaison s'explique par le fait que ces individus ayant un niveau licence ou master ont l'une des professions(H,C,A,F,I, D) et travaillent dans une petite entreprise avec comme secteur d'activité une secteur mixte ou ONG aussi il sont des travailleurs indépendant.

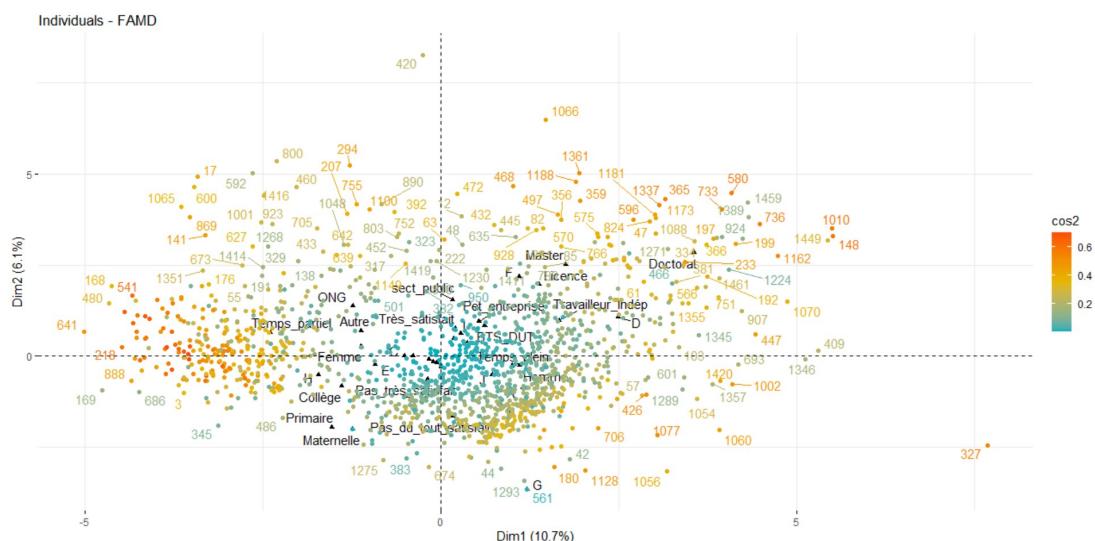


Figure 3.9 – Graphique des individus Allemagne

**Interprétation :** Comme dans le cas de la France on constate que presque tous les individus sont également proche les uns des autres proches. On voit que les individus 1271, 334, 466, 381, sont proche du niveau d'étude doctorat ce qui veut dire que ces individus ont le niveau doctorat. Aussi on constate que les individus en vert sont proche des variables tel que le niveau d'étude la satisfaction le type de contrat, la profession, le secteur(secteur public, ONG), la taille de l'entreprise(petite entreprise). Cette liaison explique que ces individus ont aux moins une de ces caractéristiques(variables) cité.

## CHAPITRE 4

---

### RÉGRESSION PAR LES MCO DES DÉTERMINANTS DES SALAIRES ET DES DIFFÉRENCES DES 2 PAYS.

#### 4.1 Objectif.

L'objectif de la régression par les MCO a pour but de tester la significativité des variables expliquant le mieux le salaire des individus. Pour cela nous avons faire une régression par les MCO de toutes les variables et une régression utilisant l'équation de gain de Mincer avec pour variables explicatives le taux horaire. Ensuite nous avons fait un certains nombre de test qui sont : le test de Chow pour voir si les hommes et les femmes ont le même salaire ou pas, le test de normalité pour tester la normalité des résidus et le test d'homoscédasticité pour vérifie si les erreurs sont homoscédastiques ou pas. Nous avons fait d'autre regression sont dans le code.  
**NB :** Dans toute notre étude nous allons considérer notre seuil de significativité à 5%.

#### 4.2 Première régression par la MCO de tous les individus des deux pays.

Après la régression par les MCO on obtient le résultat présenté dans le tableau ci-dessous.

## Chapitre 4. Régression par les MCO des déterminants des salaires et des différences entre les 2 pays.

---

```

Call:
lm(formula = taux_horaires ~ genre + age + situation + secteur_trav +
ancienneté + satisfaction + nivx_etude + profession_rec +
heures_mois + Taille_entreprise, data = ewcs_2015_fra)

Residuals:
    Min      1Q  Median      3Q     Max 
-17.662 - 3.046 - 0.677  1.992 130.607 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 15.755864   3.064874  5.141 3.18e-07 ***  
genreHomme    0.986475   0.455291  2.167 0.030449 *    
age          0.092682   0.023013  4.027 5.98e-05 ***  
situationTravailleur_Indép 0.568542   1.035686  0.549 0.583138    
secteur_traventre_mixte  5.366491   2.887934  1.858 0.063372 .    
secteur_travONG       -0.224866   2.916676 -0.077 0.938559    
secteur_travsect_privé 1.539802   2.557382  0.602 0.547217    
secteur_travsect_public 0.577400   2.592947  0.223 0.823820    
ancienneté        0.083468   0.025049  3.332 0.000887 ***  
satisfactionPas_très_satisfait 0.499451   1.063851  0.469 0.638814    
satisfactionSatisfait      1.487878   0.982169  1.515 0.130056    
satisfactionTrès_satisfait 0.820241   1.051189  0.780 0.435365    
nivx_etudecollège       -2.256016   1.030267 -2.190 0.028730 *    
nivx_etudeLicence        0.756900   0.838272  0.903 0.366740    
nivx_etudeLycée         -2.265967   0.621997 -3.643 0.000281 ***  
nivx_etudeMaster        2.605561   0.773743  3.367 0.000782 ***  
nivx_etudePrimaire      -4.893459   1.557279 -3.142 0.001716 **  
profession_recB        -0.530593   1.030920 -0.515 0.606870    
profession_recC        0.661905   1.029798  0.643 0.520505    
profession_recD        4.614862   1.248971  3.695 0.000230 ***  
profession_recE        -1.963398   0.907195 -2.164 0.030636 *    
profession_recF        2.668535   1.049764  2.542 0.011142 *    
profession_recG        5.248527   2.826899  1.857 0.063601 .    
profession_recH        -2.137211   0.964600 -2.216 0.026898 *    
profession_recI        1.025382   0.930311  1.102 0.270593    
profession_recJ        0.624209   1.981817  0.315 0.752839    
heures_mois        -0.062640   0.005404 -11.592 < 2e-16 ***  
Taille_entreprisemoy_entreprise -1.736649   0.474802 -3.658 0.000265 ***  
Taille_entreprisePet_entreprise -2.935539   0.963294 -3.047 0.002357 **  
Taille_entreprisePet_moy_entreprise -1.618916   0.670647 -2.414 0.015925 *  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 7.073 on 1234 degrees of freedom
Multiple R-squared:  0.2905, Adjusted R-squared:  0.2739 
F-statistic: 17.43 on 29 and 1234 DF,  p-value: < 2.2e-16

```

Résultat de la régression (France)

**Interprétation(France) :** Le résultat ci-dessus est le résultat de la régression par les MCO de tous les individus utilisant toutes les variables de la base de données de la France. Dans cette partie on constate que l'âge, l'ancienneté, la taille de l'entreprise, le nombre d'heure par mois, le niveau d'étude (lycée, collège, Master, Primaire), la profession(D, E, F, H) sont les variables qui expliquent le mieux le salaire car leur p-value après régression est inférieur au seuil de significativité qui est de 5% soit 0.05. Quand aux autres variables elles n'ont aucun effet sur le salaire car leur p-value est supérieur au seuil de significativité. En effet, la non significativité i.e (la non influence sur le salaire) du niveau d'étude licence s'explique par le fait que peu d'individu travail avec un niveau, car tout le monde possède une licence qui est un niveau qui n'est pas beaucoup demandé sur le marché de l'emploi. Par la suite, nous allons faire la régression par les MCO de la base de données de l'Allemagne et on obtient les résultats suivants :

## Chapitre 4. Régression par les MCO des déterminants des salaires et des différences entre les 2 pays.

---

```

call:
lm(formula = salaire ~ genre + age + situation + secteur_trav +
ancienneté + satisfaction + nivx_etude + profession_rec +
heures_mois + Taille_entreprise, data = ewcs_2015_ger)

Residuals:
    Min      1Q  Median      3Q     Max 
-2354.2 -330.0   -66.2   233.0  3635.2 

Coefficients:
                                         Estimate Std. Error t value Pr(>|t|)    
(Intercept)                         45.4566   269.6132  0.169 0.866136    
genreHomme                           291.5664   38.4697  7.579 6.21e-14 ***  
age                                  -0.6772   1.5727 -0.431 0.666833    
situationTravailleur_Indép          432.9692   76.8589  5.633 2.12e-08 ***  
secteur_traventre_mixte             190.1391   180.3235  1.054 0.291862    
secteur_travONG                      -41.4753   194.8053 -0.213 0.831430    
secteur_travsect_privé              189.1447   170.8998  1.107 0.268584    
secteur_travsect_public              203.7369   176.7140  1.153 0.249136    
ancienneté                            13.8706   2.0328  6.824 1.31e-11 ***  
satisfactionPas_très_satisfait       101.7572   147.3014  0.691 0.489797    
satisfactionSatisfait                268.8986   141.2590  1.904 0.057164 .    
satisfactionTrès_satisfait          421.4332   143.6370  2.934 0.003399 **  
nivx_etudeCollège                   -438.9405   126.8197 -3.461 0.000554 ***  
nivx_etudeDoctorat                  857.4356   243.7628  3.517 0.000449 ***  
nivx_etudeLicence                   227.3598   123.1169  1.847 0.064997 .    
nivx_etudeLycée                     -296.8546   105.7118 -2.808 0.005050 **  
nivx_etudeMaster                    274.1217   126.0375  2.175 0.029799 *    
nivx_etudeMaternelle                -357.9059   617.7296 -0.579 0.562418    
nivx_etudePost_secondaire           -135.8760   117.4788 -1.157 0.247628    
nivx_etudePrimaire                  -727.2785   206.9571 -3.514 0.000455 ***  
profession_recB                     -129.0601   68.5103 -1.884 0.059793 .    
profession_recC                     -22.0277   65.6384 -0.336 0.737228    
profession_recD                     284.8963   133.6327  2.132 0.033182 *    
profession_recE                     -97.5942   59.4023 -1.643 0.100615    
profession_recF                     161.0936   76.8707  2.096 0.036289 *    
profession_recG                     -64.4583   626.9688 -0.103 0.918129    
profession_recH                     -72.1766   74.3315 -0.971 0.331707    
profession_recI                     77.8391   66.1288  1.177 0.239358    
profession_recJ                     6.2968    146.4384  0.043 0.965708    
heures_mois                          8.7965    0.4146  21.216 < 2e-16 ***  
Taille_entrepriseMoy_entreprise    -130.1103   39.5743 -3.288 0.001034 **  
Taille_entreprisePet_entreprise    -556.8604   104.8074 -5.313 1.25e-07 ***  
Taille_entreprisePet_moy_entreprise -197.5055   47.2356 -4.181 3.07e-05 ***  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 605 on 1434 degrees of freedom
Multiple R-squared:  0.5593,    Adjusted R-squared:  0.5494 
F-statistic: 56.86 on 32 and 1434 DF,  p-value: < 2.2e-16

```

Résultat de la régression (Allemagne)

### Interprétation(Allemagne) :

D'après les résultats, on constate que les variables telle que la situation (travailleurs indépendant(e)s), l'ancienneté(e), le niveau de satisfaction (très satisfait), le niveau d'étude (collège, doctorat, Lycée, Master, primaire), la profession (D & F), le nombre d'heure de travail par mois et la taille de l'entreprise sont les variables qui influencent le plus le salaire à cause de leur p-value. En effet, ces variables expliquent le mieux le salaire, comme on peut le voir pour l'ancienneté plus un individu à plus d'expérience plus son salaire évolu.

### 4.3 Deuxième régression par les MCO utilisant l'équation de gain de Mincer de tous les individus.

Notre objectif ici est d'utiliser l'équation de gain de Mincer pour effectuer cette régression mais pour cela nous utiliserons comme variable explicative le taux horaires en lieu et place du salaire vue que nous avons déjà utilisé le salaire comme variable explicative dans la première régression et comme variable à expliquer le niveau d'étude et l'ancienneté de l'individu. Comme nous l'avons présenté plus haut l'équation de Mincer se définir comme suit :

$$\ln w = \ln w_0 + \rho s + \beta_1 x + \beta_2 x^2$$

En utilisant cette formule pour notre régression par les MCO on obtient le résultat suivant :

#### Analysis of Variance Table

```
Response: log(taux_horaires)
          Df  Sum Sq Mean Sq F value Pr(>F)
nivx_etude      5   43.703  8.7407 60.1583 <2e-16 ***
ancienneté      1   19.798 19.7983 136.2631 <2e-16 ***
I(ancienneté^2) 1   0.067  0.0666  0.4585 0.4985
Residuals       1256 182.490  0.1453
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Résultat de la régression utilisant l'équation de Mincer (France)

On remarque bien que le niveau d'étude et l'ancienneté sont bien significatives car p-value inférieure à 5%. Par conséquent le niveau d'études et l'ancienneté expliquent le taux horaire donc le salaire.

#### Analysis of Variance Table

```
Response: log(taux_horaires)
          Df  Sum Sq Mean Sq F value Pr(>F)
nivx_etude      8   32.956  4.1195 29.5756 < 2e-16 ***
ancienneté      1   13.035 13.0352 93.5845 < 2e-16 ***
I(ancienneté^2) 1   0.762  0.7621  5.4711 0.01947 *
Residuals       1456 202.803  0.1393
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Résultat de la régression utilisant l'équation de Mincer (Allemagne)

Au vue de ces résultats on constate bien que le niveau d'études et l'ancienneté sont significatif mais l'ancienneté<sup>2</sup> n'est pas significatif. Par conséquent seul le niveau d'études et l'ancienneté expliquent mieux le taux horaire donc le salaire. Au vue de ces deux résultats nous pouvons dire que le niveaux d'étude et l'ancienneté de l'individu sont des déterminant du taux horaire donc du salaire.

**NB :** Pour cette régression nous n'avons pas fait de test.

## 4.4 Les différents types de tests de validité du modèle associé à la première régression

### 4.4.1 Test de Chow

L'objectif de cette partie est de voir entre l'homme et la femme celui qui touche plus de salaire mensuelles. Pour ce test, nous allons utiliser le test de Chow ayant les hypothèses suivantes :

- $H_0$  : les hommes et les femmes touche les même revenu(salaire) i.e a dire p-value > 5%.
- $H_1$  : Le cas contraire p-value < 5%

France		Allemagne	
Stat F du test de Chow	p-value	Stat F du test de Chow	p-value
2.796	0.000	-23.516	1.000

Tableau 4.1 – Résultat du test de Chow

D'après le tableau ci dessus on constate que la p-value est inférieure à 5% donc les hommes et les femmes non pas la même rémunération (salaire) mensuelle. Ce résultat confirme nos résultats de visualisation quelques soit la catégorie mis en relation avec le salaire pour la base de données de la France. Au niveau de l'Allemagne on accepte  $H_0$  ce qui veut dire que les Hommes et les Femmes touche les mêmes revenus cela confirme notre hypothèse de la visualisation.

### 4.4.2 Test de normalité

L'objectif du test de normalité est de faire un test sur les résidus dans le but de voir et de vérifier si les résidu obtenu à partir de notre modèle suivent une loi normale. Pour cela nous avons utiliser le test de Jacque Bera(JB) avec pour hypothèse :

- $H_0$  : Hypothèse de normalité i.e p-value > 5%
- $H_1$  : Hypothèse de non normalité.

Après estimation des résidus des deux pays on obtient les résultats suivant :

France		Allemagne	
JB	p-value	JB	p-value
787.56	<2.2e-16	44787	< 2.2e-16

Tableau 4.2 – Résultat du test de normalité

A partir de ces résultat, et quelque soit le pays on constate que les résidus ne suivent pas une loi normale ce qui veut dire que les résidus ne sont pas gaussiens. Cela se confirme par les graphiques suivant :

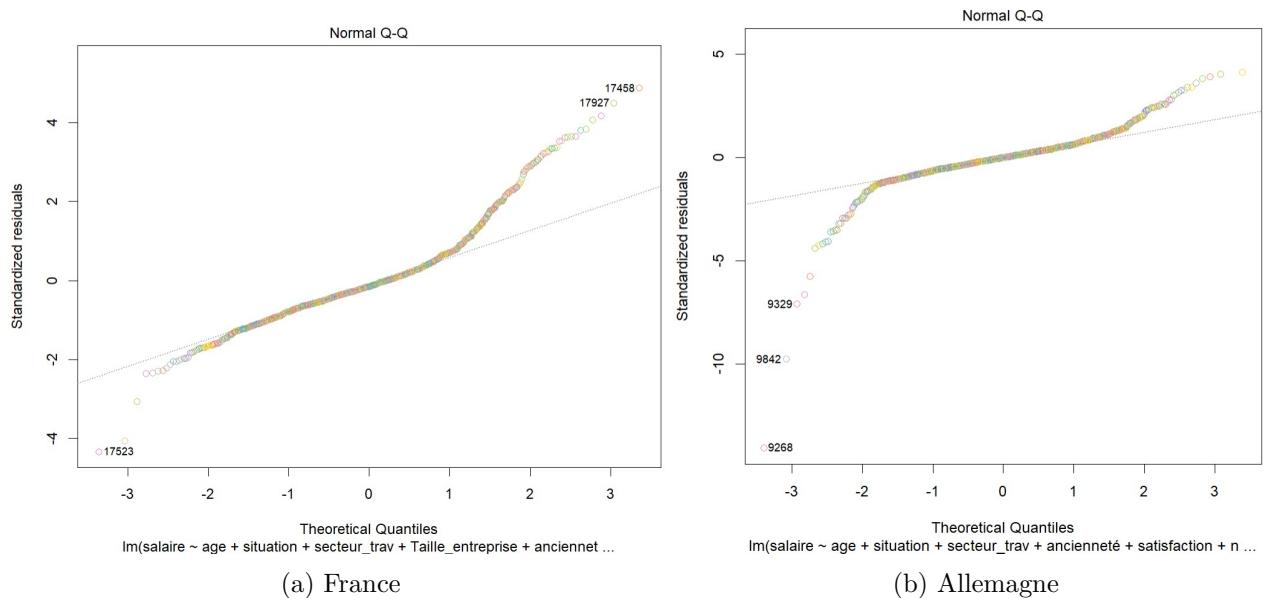


Figure 4.1 – Graphiques Test de normalité de la variable consommation (QQ-plot) qqplot

Si les résidus étaient gaussien l courbe en couleur et la droite seront proche et se suivront tout au long du parcours ce qui n'est pas totalement le cas.

#### 4.4.3 Test d'homoscédasticité vs hétéroscélasticité

L'objectif du test d'homoscédasticité est de vérifier si les erreurs sont homoscédastiques ou hétéroscédasticité. Pour ce test nous allons utiliser le test d'hétéroscédasticité de White. Les hypothèses sont les suivantes :

- $H_0$  : les erreurs sont homoscédasticité
- $H_1$  : les erreurs sont hétéroscédasticité

France		Allemagne	
Stat.LM	p-value	Stat.LM	p-value
120.2042	0.0000	851.11	0.0000

Tableau 4.3 – Résultat du test de white

On constate que quelque soit la base de données la p-value est inférieure à 5% donc on rejette  $H_0$  ce qui revient à dire que les erreurs sont hétéroscédasticité.

## CONCLUSION

Ce projet avait pour but de nous intéresser aux déterminants du salaire en France et en Allemagne à partir d'une fonction de type Mincer. On a estimé les différences salariales entre ces deux pays en s'inspirant de la méthodologie de Meurs et Ponthieux sur l'écart salariale entre les hommes et les femmes. Nous avons effectué une visualisation, une analyse et une interprétation ainsi que des régressions par les MCO des déterminants des salaires et des différences entre ces deux pays avec les données issues de l'enquête sur les individus et leurs travailles.

Nous nous sommes intéressées aux variables suivante : Le genre (Femme & Homme), l'âge, le salaire (Le revenu mensuel, le taux horaires), la taille de l'entreprise, la profession de l'individu, l'ancienneté (Nombre d'année dans l'entre-prise), le nombre d'heure de travail par semaine et par mois, le taux horaires, le type de contrat (temps plein & temps partiel), la situation de l'individu (employée ou travailleurs indépendant), le secteur du travail (Privé, Public ou ONG), le niveaux d'étude, le niveau de satisfaction de l'individu, la profession de l'individu et voir lesquels de ces variables expliquent le salaire.

Tout d'abord la visualisation des données de ces deux pays avec un niveau de développement comparable. On a observé qu'en moyenne les hommes sont mieux rémunérés que les femmes dans notre base de données cette écart est d'environ 21% en France contre 52% en Allemagne. Dans nos données nous avons plus de femmes que d'hommes en France et le contraire en Allemagne mais ces écarts ne sont pas très significatifs. On a observé que le salaire moyen en France est de 1832€/mois contre 1683€/mois en Allemagne avec une forte concentration des salaires autour de la moyenne dans les 2 pays. On observe que le niveau d'étude majoritaire est le niveau lycée il représente environ 50% de nos individus en France et 70% en Allemagne ce qui pourrait fortement biaiser les résultats car l'échantillon n'est pas assez grand et ne pourrait pas refléter réellement les salaires pour les autres niveaux d'études qui sont très faiblement représentés dans cette enquête. On a observés que en moyenne les femmes sont plus en temps partiels que les hommes en effets en France 30% et 47% en Allemagne sont à temps partiels contre 9% des Hommes en France et 12% des Hommes en Allemagne sont à temps partiels on a observé aussi que le type de contrat a un impact sur le salaire en effet en Allemagne un individus en temps plein touche 106% de plus que un individus en temps partiels et en France cette écart est d'environ 64%. On observe aussi que la taille de l'entreprise joue un rôle sur le salaire en France plus l'entreprise est grande plus le salaire est élevé en Allemagne c'est aussi le cas mais très légèrement. On observe que les années d'expériences ont aussi un impact sur le salaire, plus on

## **Chapitre 4. Régression par les MCO des déterminants des salaires et des différences entre les 2 pays.**

---

a des années d'expériences plus le salaire augmente, le secteur d'activité a aussi un impact les secteurs d'activité ne se valent pas tous certains secteurs sont mieux rémunérée que d'autres.

L'analyse de donnée cette analyse explique faiblement notre base de données nous avons choisi 2 axes. Nous avons analysé les corrélations entre les différentes variables avec une analyse factorielle des données mixtes. Nous avons ensuite fait des régressions par les MCO.

Les régressions par les MCO ont pour objectif d'expliquer le salaire en fonctions d'autres variables qui ont un impact sur le salaire et de mesurer cette impact nous avons fait une régression avec toutes les variables expliquant le salaire nous avons observés que l'âge, l'ancienneté, la taille de l'entreprise, le nombre d'heure par mois, le niveau d'étude (lycée, collège, Master, Primaire), la profession(D, E, F, H) sont les variables qui expliquent le mieux le salaire en France , en Allemagne la situation(travailleur indépendant(e)s), l'ancienneté(e), le niveau de satisfaction(très satisfait), le niveau d'étude(collège, doctorat, Lycée, Master, primaire), la profession(D & F), le nombre d'heure de travail par mois et la taille de l'entreprise sont les variables qui expliquent le mieux le salaire . Nous avons aussi fait les régressions par les MCO avec la fonction de type Mincer on a conclu que les années d'expériences et le niveau d'étude détermine le salaire mais ne l'explique pas à 100% il y a d'autres variable qui ont un impact sur le salaire. Certains résultats dans nos régressions sont biaisés cela est dû aux données qui ne sont peut-être pas fiable car cela reste un questionnaire et les individus sont libres de leurs réponses et à la taille de l'échantillon qui reste assez faibles si l'on prend que ces 2 pays.

## ANNEXES A

---

### A.1 Annexe 1

Catégories	France		Allemagne	
	Homme %	Femme %	Homme %	Femme %
<b>Niveau d'étude</b>				
Doctorat	0.00	0.00	0.41	0.69
Master	16.78	18.46	6.49	4.95
Licence	7.17	11.08	7.70	3.99
DUT_BTS	16.78	16.00	2.70	2.06
Lycée	50.81	46.92	71.22	70.56
Collège	6.68	5.08	3.92	6.74
Post_secondaire	0.00	0.00	6.89	9.90
Primaire	1.79	2.46	0.68	0.96
Maternelle	0.00	0.00	0.00	0.14
<b>Sous total en %</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>
<b>Type de contrat</b>				
Temps plein	91.21	69.69	87.97	52.96
Temps partiel	8.79	30.31	12.03	47.04
<b>Sous total en %</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>
<b>Secteur d'activité</b>				
Secteur privée	71.82	62.00	81.22	74.83
secteur public	23.13	32.92	9.59	12.65
Secteur mixte	2.93	1.69	7.70	6.60
Autre	0.65	0.62	0.27	1.51
ONG	1.47	2.77	1.22	4.40
<b>Sous total en %</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>

Tableau A.1 – Répartition en pourcentage du genre en fonction des autres variables

## Annexes A.

---

### A.2 Annexe 2

	France									
	A	B	C	D	E	F	G	H	I	J
Femmes %	0.77	1.38	16.00	4.15	26.31	18.15	0.15	14.46	17.38	1.23
Hommes %	16.29	13.19	4.72	6.35	10.10	19.22	0.98	8.79	19.06	1.30

Tableau A.2 – Répartition en pourcentage du genre en fonction de la profession

### A.3 Annexe 3

	France									
	A	B	C	D	E	F	G	H	I	J
Femmes %	1.65	2.89	19.39	1.38	32.32	11.14	0.00	12.38	18.29	0.55
Hommes %	25.68	15.81	9.05	2.03	15.27	12.03	0.14	6.89	11.08	2.03

Tableau A.3 – Répartition en pourcentage du genre en fonction de la profession

### A.4 Annexe 4

Variables	Min	1 <sup>er</sup> Q	Médiane	Mean	3 <sup>rd</sup> Q	Max	Écart-type
<b>Niveau d'étude</b>							
Master	300	2000	2600	2832	3550	5500	1174.412
Licence	300	1525	2000	2055	2400	4500	898.6873
DUT_BTS	400	1575	2000	2097	2500	5000	770.4349
Lycée	300	1300	1600	1772	2000	5000	726.1346
Collège	300	1312	1500	1692	1968	4500	782.9352
Primaire	218	1375	1500	1449	1710	2500	604.8315
<b>Type de contrat</b>							
Temps plein	300	1500	1850	2094	2500	5500	901.4827
Temps partiel	218	700	900	1175	1500	3500	730.2809
<b>Secteur</b>							
Secteur privée	218	1350	1700	1935	2250	5000	894.2734
Secteur public	570	1600	2000	2212	2600	5200	938.5375
Secteur mixte	780	2000	2200	2671	2975	5500	1239.814
Autre	750	1088	1400	1488	1800	2400	700.4463
ONG	1112	1200	1610	1597	1900	2300	456.0131
<b>Situation</b>							
Employé(e)s	218	1400	1800	2031	2400	5500	923.6868
Travailleurs indépendants	300	1212	1750	1812	2000	4500	920.6427

Tableau A.4 – Statistique descriptive des variables en fonction du salaire

## A.5 Annexe 5

Variables	Min	1 <sup>er</sup> Q	Médian	Femmes (France)			
				Mean	3 <sup>rd</sup> Q	Max	Écart-type
<b>Niveau d'étude</b>							
Master	1300	1760	2350	2482	3000	5000	1095.028
Licence	250	1500	1832	1921	2125	5000	908.3787
DUT_BTS	160	1300	1600	1823	2000	5000	904.5133
Lycée	189	1023	1300	1310	1500	4000	529.5341
Collège	450	750	1115	1185	1500	3200	553.3529
Primaire	300	575	860	1018	1221	3000	680.0008
<b>Type de contrat</b>							
Temps partiel	136	700	1023	1220	1600	4500	764.6881
Temps plein	350	1300	1550	1854	2100	5000	887.26
<b>Secteur</b>							
Secteur privée	136	1100	1360	1595	1900	5000	908.4109
Secteur public	203	1300	1654	1796	2084	5000	869.2227
Secteur mixte	750	1326	1637	1999	2000	5000	1303.847
Autre	634.0	983.5	1150.0	1108.5	1275.0	1500.0	359.104
ONG	386	1150	1450	1484	1800	3000	639.1392
<b>Situation</b>							
Employé(e)s	136	1130	1450	1626	1950	5000	866.4482
Travailleuses indépendantes	350	1500	2000	2337	3000	5000	1217.404

Tableau A.5 – Statistique descriptive des variables en fonction du salaire

## A.6 Annexe 6

Variables	Homme (Allemagne)						
	Min	1 <sup>er</sup> Q	Médian	Mean	3 <sup>rd</sup> Q	Max	Ecart type
<b>Niveau d'étude</b>							
Doctorat	2500	3250	4000	3833	4500	5000	1258.306
Master	450	1800	2400	2769	4000	5600	1379.002
Licence	300	2100	2700	2723	3200	5800	897.2637
DUT_BTS	1200	1975	2300	2322	2688	4000	673.6536
Lycée	130	1480	1800	1894	2300	5200	752.9929
Collège	130	700	1500	1298	1730	2800	712.833
Post_secondaire	450	1782	2200	2210	2550	4500	844.6137
Primaire	332	1200	1250	1086	1250	1400	428.3396
<b>Type de contrat</b>							
Temps plein	200	1580	2000	2134	2500	5800	823.0492
Temps partiel	130	450	900	1239	1800	5600	1003.625
<b>Secteur</b>							
Secteur privée	130	1500	1900	2023	2500	5800	895.9902
secteur public	320	1600	2000	2186	2650	4000	818.5292
Secteur mixte	130	1500	1900	1998	2500	4500	926.6644
Autre	1150	1488	1825	1825	2162	2500	954.5942
ONG	200	600	850	1214	1800	2700	882.328

Tableau A.6 – Statistique descriptive des variables en fonction du salaire

## A.7 Annexe 7

Variables	Femme (Allemagne)						
	Min	1 <sup>er</sup> Q	Median	Mean	3 <sup>rd</sup> Q	Max	Ecart type
<b>Niveau d'étude</b>							
Doctorat	1750	3000	3500	3310	3500	4800	1097.953
Master	144	1688	2350	2263	3000	3800	898.2495
Licence	400	1200	1800	1922	2600	5000	1179.031
DUT_BTS	180	1100	1700	1625	2125	3400	896.0538
Lycée	120	800	1185	1215	1550	4500	629.3038
Post_secondaire	250	1015	1400	1547	1862	4000	748.5407
Collège	200	639	930	1030	1250	3200	604.2635
Primaire	350.0	550.0	765.0	780.7	1025.0	1200.0	321.0974
Maternelle	1100	1100	1100	1100	1100	1100	
<b>Type de contrat</b>							
Temps partiel	100	121	100	121	121	100	121
Temps partiel	100	121	100	121	121	100	121
<b>Secteur</b>							
Secteur privée	100	121	100	121	121	100	121
secteur public	100	121	100	121	121	100	121
Secteur mixte	100	121	100	121	121	100	121
Autre	100	121	100	121	121	100	121
ONG	100	121	100	121	121	100	121

Tableau A.7 – Statistique descriptive des variables en fonction du salaire

## A.8 Annexe 8

	France		Allemagne	
	Valeurs propres	Variance en %	Valeurs propres	Variance en %
Dim. 1	3.29	9.96	3.86	10.71
Dim. 2	2.45	7.43	2.19	6.08
Dim. 3	1.82	5.53	1.84	5.10
Dim. 4	1.73	5.25	1.57	4.35
Dim. 5	1.45	4.38	1.42	3.95

Tableau A.8 – Valeurs propres et variance cumulée

## A.9 Annexes 9

Variables	France		Allemagne	
	Dim. 1	Dim. 2	Dim. 1	Dim. 2
âge	0.17	-0.16	0.01	0.33
ancienneté	0.30	-0.23	0.31	0.33
heure_par_semaine	0.77	0.50	0.88	-0.32
heures_mois	0.77	0.50	0.88	-0.32
salaire	0.81	-0.33	0.83	0.30
taux_horaires	0.31	-0.64	0.26	0.61

Tableau A.9 – Coordonnées variables quantitatives

## A.10 Annexe 10

```

call:
lm(formula = taux_horaires ~ genre + age + situation + secteur_trav +
    Taille_entreprise + ancienneté + satisfaction + nivx_etude +
    profession_rec + heures_mois, data = data_temps_plein)

Residuals:
      Min        1Q     Median       3Q      Max 
-17.154   -2.617   -0.589   1.742  88.029 

Coefficients:
                                         Estimate Std. Error t value Pr(>|t|)    
(Intercept)                         15.160910  2.976977  5.093 4.23e-07 ***
genreHomme                           0.895584  0.386767  2.316 0.020787 *  
age                                    0.068920  0.020767  3.319 0.000937 *** 
situationTravailleur_Indép          1.048426  0.916205  1.144 0.252772  
secteur_traventre_mixte             8.911513  2.734656  3.259 0.001157 **  
secteur_travONG                      0.713247  2.801692  0.255 0.799103  
secteur_travsect_privé              3.494593  2.485222  1.406 0.159996  
secteur_travsect_public              2.775602  2.513358  1.104 0.269716  
Taille_entreprisemoy_entreprise    -1.351076  0.407130  -3.319 0.000938 *** 
Taille_entreprisePet_entreprise    -2.648707  0.890816  -2.973 0.003018 **  
Taille_entreprisePet_moy_entreprise -1.651302  0.595696  -2.772 0.005676 ** 
ancienneté                            0.101908  0.021728  4.690 3.11e-06 *** 
satisfactionPas_très_satisfait     0.615341  0.964307  0.638 0.523546  
satisfactionSatisfait               1.277892  0.900963  1.418 0.156402  
satisfactionTrès_satisfait         1.220296  0.960096  1.271 0.204024  
nivx_etudecollège                  -1.511577  0.908340  -1.664 0.096409 .  
nivx_etudeLicence                  -0.038349  0.716888  -0.053 0.957349  
nivx_etudeLycée                   -1.830232  0.529452  -3.457 0.000570 *** 
nivx_etudeMaster                  3.525398  0.657637  5.361 1.03e-07 *** 
nivx_etudePrimaire                -3.169383  1.470732  -2.155 0.031406 *  
profession_recB                  -0.232032  0.823882  -0.282 0.778284  
profession_recc                  -0.405156  0.855308  -0.474 0.635822  
profession_recD                  4.367913  0.995001  4.390 1.26e-05 *** 
profession_recE                  -0.956624  0.758896  -1.261 0.207771  
profession_recF                  2.793635  0.854093  3.271 0.001109 **  
profession_recG                  5.188572  2.191029  2.368 0.018072 *  
profession_recH                  -1.620487  0.820383  -1.975 0.048516 *  
profession_recI                  1.472537  0.746430  1.973 0.048801 *  
profession_recJ                  0.806140  1.709838  0.471 0.637408  
heures_mois                       -0.069482  0.005778 -12.024 < 2e-16 *** 

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.439 on 983 degrees of freedom
Multiple R-squared:  0.4006,    Adjusted R-squared:  0.383 
F-statistic: 22.66 on 29 and 983 DF,  p-value: < 2.2e-16

```

Résultats de la régression (temps plein) France

## A.11 Annexe 11

```

Call:
lm(formula = taux_horaires ~ genre + age + situation + secteur_trav +
    ancienneté + satisfaction + nivx_etude + profession_rec +
    heures_mois, data = data_temps_plein)

Residuals:
    Min      1Q  Median      3Q     Max 
-12.343 -2.514 -0.647  1.736 44.897 

Coefficients:
                                         Estimate Std. Error t value Pr(>|t|)    
(Intercept)                      21.862480   2.678943   8.161 9.87e-16 ***
genreHomme                         2.416058   0.353888   6.827 1.50e-11 ***
age                                  0.009517   0.015506   0.614  0.53950  
situationTravailleur_Indép       3.038619   0.545166   5.574 3.20e-08 ***
secteur_traventre_mixte          1.536436   1.779497   0.863  0.38812  
secteur_travONG                   -0.356435   1.974406  -0.181  0.85677  
secteur_travsect_privé           1.532116   1.705844   0.898  0.36932  
secteur_travsect_public          2.050037   1.754652   1.168  0.24294  
ancienneté                          0.106591   0.019307   5.521 4.29e-08 ***
satisfactionPas_très_satisfait  1.083047   1.353908   0.800  0.42393  
satisfactionSatisfait            2.133061   1.297634   1.644  0.10053  
satisfactionTrès_satisfait       3.226254   1.320197   2.444  0.01471 *  
nivx_etudecollège                -3.062702   1.220167  -2.510  0.01223 *  
nivx_etudeDoctorat              4.961704   1.990283   2.493  0.01283 *  
nivx_etudeLicence                1.843846   1.134686   1.625  0.10448  
nivx_etudeLycée                  -2.220055   0.983204  -2.258  0.02416 *  
nivx_etudeMaster                 1.626561   1.152690   1.411  0.15852  
nivx_etudeMaternelle             -3.633182   4.902465  -0.741  0.45881  
nivx_etudePost_secondaire        -0.471836   1.087667  -0.434  0.66452  
nivx_etudePrimaire               -5.503570   1.969216  -2.795  0.00529 ** 
profession_recB                  -0.428558   0.565431  -0.758  0.44867  
profession_recC                  -0.136950   0.578668  -0.237  0.81297  
profession_recD                  2.851800   1.139483   2.503  0.01248 *  
profession_recE                  -0.241719   0.535666  -0.451  0.65191  
profession_recF                  1.497648   0.692597   2.162  0.03083 *  
profession_recG                  0.985211   4.974712   0.198  0.84305  
profession_recH                  -0.914881   0.720351  -1.270  0.20436  
profession_recI                  0.415660   0.578278   0.719  0.47244  
profession_recJ                  -0.130488   1.249665  -0.104  0.91686  
heures_mois                        -0.094433   0.006335 -14.906 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.748 on 1006 degrees of freedom
Multiple R-squared:  0.3531,    Adjusted R-squared:  0.3345 
F-statistic: 18.94 on 29 and 1006 DF,  p-value: < 2.2e-16

```

Résultats de la régression (temps plein) Allemagne

## A.12 Annexe 12

```

Call:
lm(formula = log(salaire) ~ nivx_etude + ancienneté + I(ancienneté^2),
  data = ewcs_2015_fra)

Residuals:
    Min      1Q  Median      3Q     Max 
-2.28783 -0.20572  0.01574  0.25045  1.39603 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 7.2671260  0.0371190 195.779 < 2e-16 ***
nivx_etudeCollège -0.3096743  0.0602620 -5.139 3.20e-07 ***
nivx_etudeLicence  0.0126196  0.0516117  0.245  0.8069  
nivx_etudeLycée   -0.2418429  0.0357096 -6.772 1.94e-11 *** 
nivx_etudeMaster   0.2982607  0.0429223  6.949 5.90e-12 *** 
nivx_etudePrimaire -0.6189242  0.0913241 -6.777 1.88e-11 *** 
ancienneté        0.0250854  0.0038730  6.477 1.34e-10 *** 
I(ancienneté^2)   -0.0002789  0.0001141 -2.444  0.0147 *  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4445 on 1256 degrees of freedom
Multiple R-squared:  0.2728, Adjusted R-squared:  0.2687 
F-statistic: 67.3 on 7 and 1256 DF, p-value: < 2.2e-16

```

Résultats régression avec équation de Mincer de tous les individus (France)

## A.13 Annexe 13

```

call:
lm(formula = log(salaire) ~ nivx_etude + ancienneté + I(ancienneté^2),
  data = ewcs_2015_ger)

Residuals:
    Min      1Q  Median      3Q     Max 
-2.75232 -0.27055  0.08328  0.34907  1.23694 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 7.2106862  0.0953562 75.618 < 2e-16 ***
nivx_etudeCollège -0.5315595  0.1107669 -4.799 1.76e-06 ***
nivx_etudeDoctorat 0.5883993  0.2126408  2.767 0.005727 ** 
nivx_etudeLicence  0.1576123  0.1088329  1.448 0.147775  
nivx_etudeLycée   -0.2555237  0.0932892 -2.739 0.006237 ** 
nivx_etudeMaster   0.2147742  0.1091865  1.967 0.049368 *  
nivx_etudeMaternelle -0.4282340  0.5502240 -0.778 0.436524  
nivx_etudePost_secondeaire -0.0918417  0.1039662 -0.883 0.377176  
nivx_etudePrimaire   -0.6538087  0.1818733 -3.595 0.000335 *** 
ancienneté        0.0358306  0.0043149  8.304 2.28e-16 *** 
I(ancienneté^2)   -0.0006163  0.0001250 -4.932 9.06e-07 *** 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.5425 on 1456 degrees of freedom
Multiple R-squared:  0.1899, Adjusted R-squared:  0.1844 
F-statistic: 34.13 on 10 and 1456 DF, p-value: < 2.2e-16

```

Résultats régression avec équation de Mincer de tous les individus (Allemagne)

## Annexes A.

---

### A.14 Annexes 14

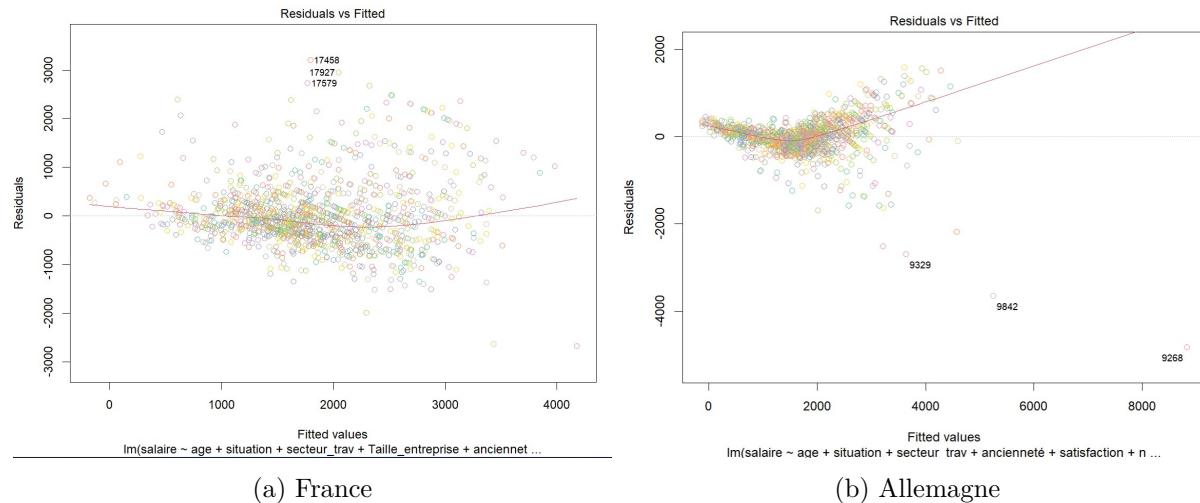


Figure A.1 – Graphiques des tests

### A.15 Annexes 15

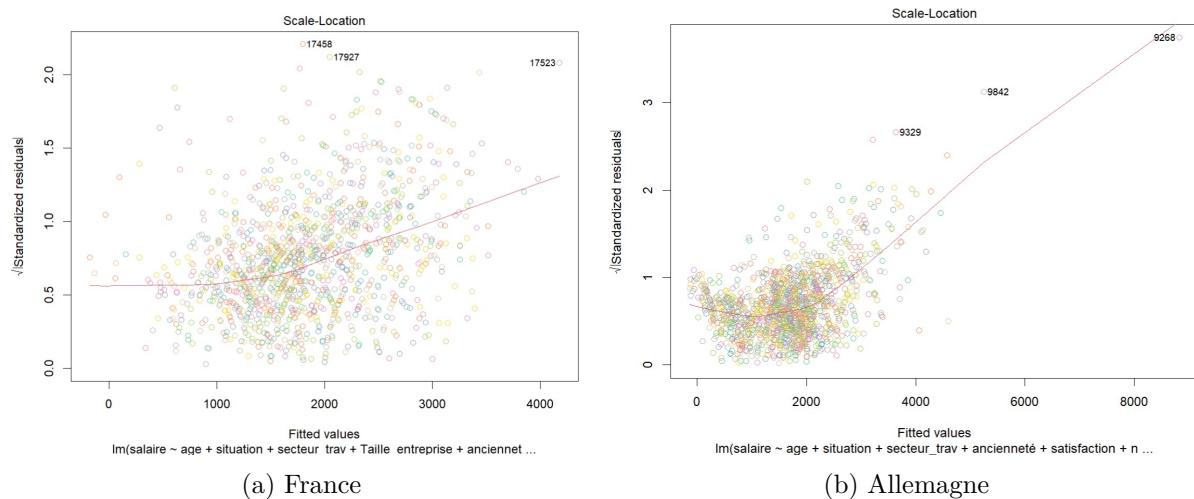


Figure A.2 – Graphiques des tests

## Annexes A.

---

### A.16 Annexe 16

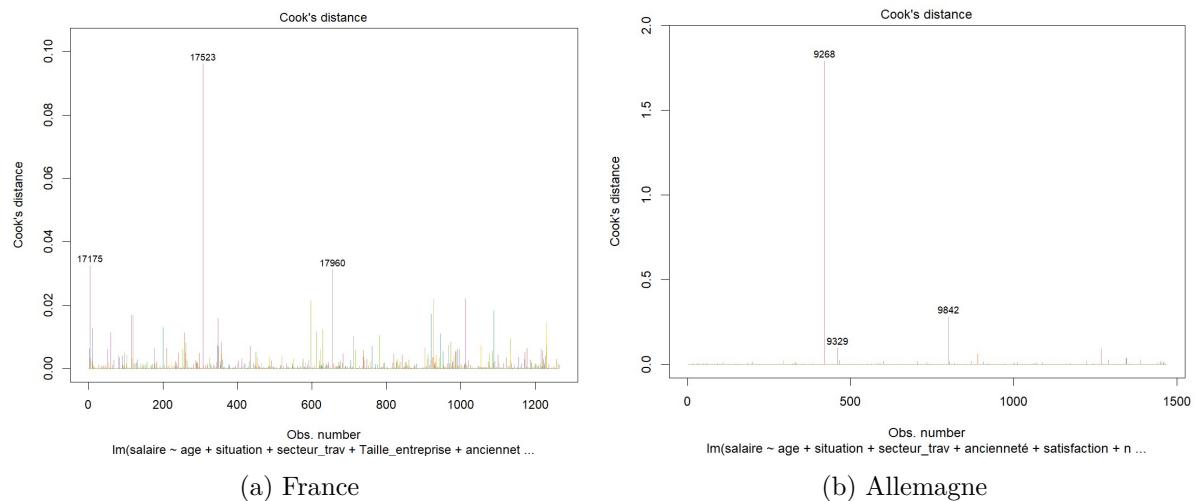


Figure A.3 – Graphiques des tests

### A.17 Annexes 17

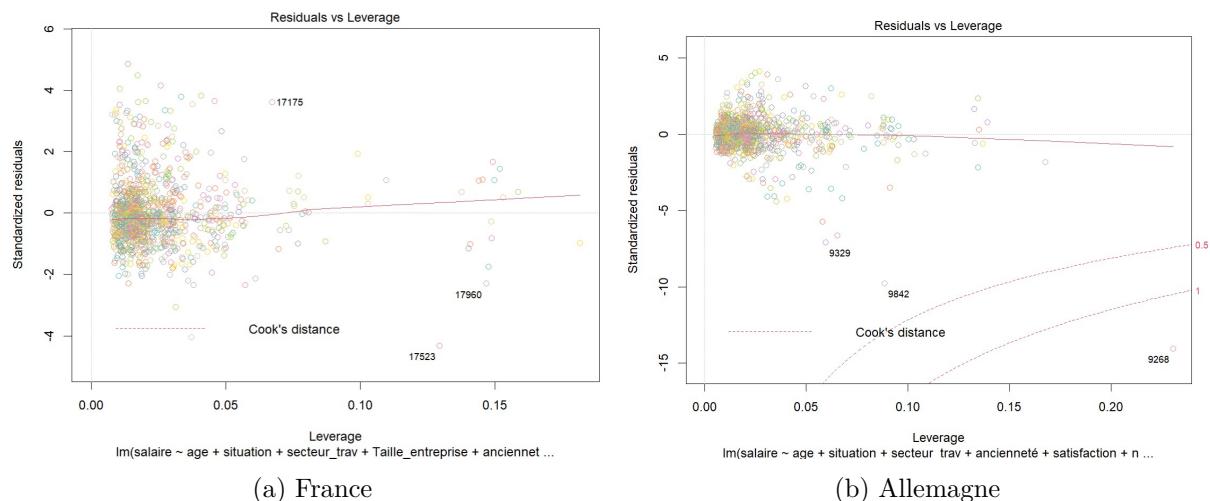


Figure A.4 – Graphiques des tests

## A.18 Annexes 18

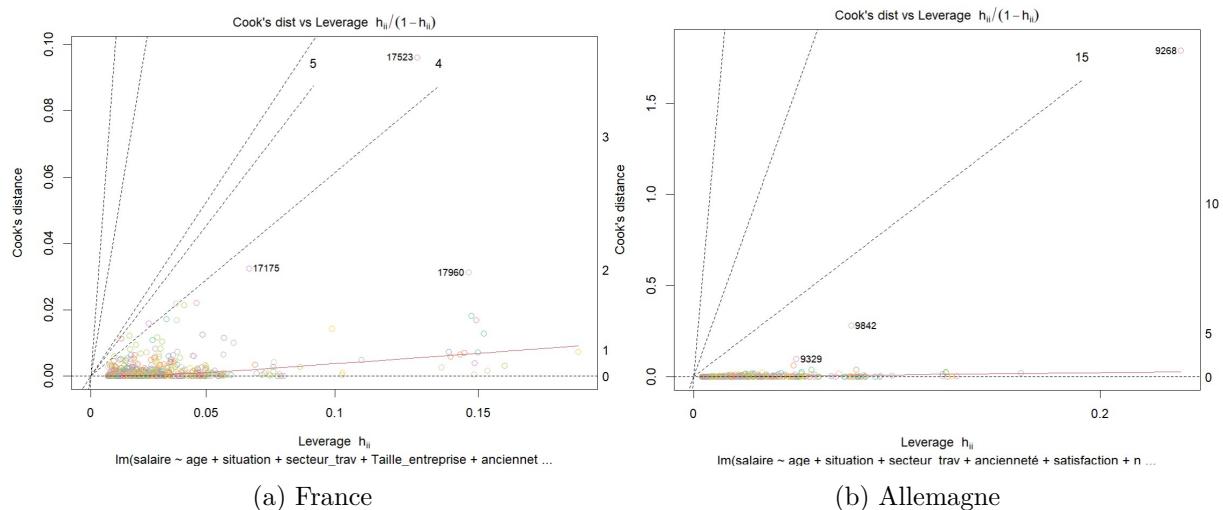


Figure A.5 – Graphiques des tests

## CODE SCRIPT DU PROJET

```
##### Projet introduction à R #####
#####
##### installation des packages #####
#####
#install.packages("factoextra")
#install.packages("normtest")
#install.packages("FactoMineR")
#install.packages("ggplot2")
#install.packages("lmtest")
#install.packages("readtext")
#install.packages("ggcorrplot")
#install.packages("readxl")
#install.packages("carData")
#install.packages("car")

#####* #### Nettoyer la "mémoire vive" de R: supprimer tous les objets créés
rm(list = ls()) # supprimer tous les objets
ls() #afficher tous les objets qui ont été créés

#####
##### les librairies #####
#####
library(lmtest)
library(ggplot2)
library(normtest)
library(factoextra)
library(FactoMineR)
library(car)
library(dplyr,quietly = T)
library(questionr)
library(data.table)
library(FactoMineR)
```

## Annexes A.

---

```
library(factoextra)
library(corrplot)
library(ggcorrplot)
library(RColorBrewer)
library(RColorBrewer)

#####
##### Importation de la base de données #####
#####

##### Working Directory (répertoire de travail)
setwd("C:/Users/fofna/OneDrive/Bureau/Cour_2021_2022/EWCS_2015")

getwd()
#


ewcs_2015 <- read.csv("ewcs_2015.csv",header = T, sep=",")
attach(ewcs_2015)

##### Choix des variables pour notre études #####
ewcs_2015_new = subset(ewcs_2015,select=c(Country,Q2a,Q2b,Q2d,Q7,Q14,Q16b,Q17,Q24,Q88,Q104_€))

#### Ligne de code ligné aux choix de couleur
coul <- brewer.pal(8, "Set2")

#####
##### Pretraitement de la base de donnée #####
#####



## Recodage de la variable Q2a
ewcs_2015_new$Q2a <- ewcs_2015_new$Q2a
ewcs_2015_new$Q2a[ewcs_2015_new$Q2a == "Don't know (spontaneous)"] <- NA
ewcs_2015_new$Q2a <- factor(ewcs_2015_new$Q2a)

## Recodage de la variable Q2b
ewcs_2015_new$Q2b <- ewcs_2015_new$Q2b
ewcs_2015_new$Q2b[ewcs_2015_new$Q2b == "DK (spontaneous)"] <- NA
ewcs_2015_new$Q2b[ewcs_2015_new$Q2b == "Refusal (spontaneous)"] <- NA
ewcs_2015_new$Q2b <- as.numeric(ewcs_2015_new$Q2b)

## Recodage de ewcs_2015_new$Q16b en ewcs_2015_new$Q16b_rec
ewcs_2015_new$Q16b <- ewcs_2015_new$Q16b
ewcs_2015_new$Q16b[ewcs_2015_new$Q16b == "1 (interviewee works alone)"] <- "Pet_entreprise"
ewcs_2015_new$Q16b[ewcs_2015_new$Q16b == "10-249"] <- "moy_entreprise"
ewcs_2015_new$Q16b[ewcs_2015_new$Q16b == "2-9"] <- "Pet_moy_entreprise"
ewcs_2015_new$Q16b[ewcs_2015_new$Q16b == "250+"] <- "Grde_entreprise"
ewcs_2015_new$Q16b[ewcs_2015_new$Q16b == "DK (spontaneous)"] <- NA
ewcs_2015_new$Q16b[ewcs_2015_new$Q16b == "Refusal (spontaneous)"] <- NA
```

## Annexes A.

---

```
ewcs_2015_new$Q16b <- factor(ewcs_2015_new$Q16b)

## Recodage de la variable Q2d
ewcs_2015_new$Q2d[ewcs_2015_new$Q2d == "DK (spontaneous)"] <- NA
ewcs_2015_new$Q2d[ewcs_2015_new$Q2d == "Refusal (spontaneous)"] <- NA
ewcs_2015_new$Q2d <- factor(ewcs_2015_new$Q2d)## Recodage de ewcs_2015_new$Q2d

## Recodage de la variable isco_08_1
ewcs_2015_new$isco_08_1[ewcs_2015_new$isco_08_1 == ""] <- NA
ewcs_2015_new$isco_08_1 <- factor(ewcs_2015_new$isco_08_1)

## Recodage de la variable Q104_euro
ewcs_2015_new$Q104_euro <- ewcs_2015_new$Q104_euro
ewcs_2015_new$Q104_euro[ewcs_2015_new$Q104_euro == "Refusal (spontaneous)"] <- NA
ewcs_2015_new$Q104_euro[ewcs_2015_new$Q104_euro == "DK (spontaneous)"] <- NA
ewcs_2015_new$Q104_euro <- as.numeric(ewcs_2015_new$Q104_euro)

## Recodage de la variable Q7
ewcs_2015_new$Q7 <- ewcs_2015_new$Q7
ewcs_2015_new$Q7[ewcs_2015_new$Q7 == "Don't know (spontaneous)"] <- NA
ewcs_2015_new$Q7[ewcs_2015_new$Q7 == "Refusal (spontaneous)"] <- NA
ewcs_2015_new$Q7 <- factor(ewcs_2015_new$Q7)

## Recodage de la variable Q14
ewcs_2015_new$Q14 <- ewcs_2015_new$Q14
ewcs_2015_new$Q14[ewcs_2015_new$Q14 == "DK (spontaneous)"] <- NA
ewcs_2015_new$Q14[ewcs_2015_new$Q14 == "Refusal (spontaneous)"] <- NA
ewcs_2015_new$Q14 <- factor(ewcs_2015_new$Q14)

## Recodage de la variable Country (pays)
ewcs_2015_new$Country <- factor(ewcs_2015_new$Country)#irec(ewcs_2015_new, "Q17")

## Recodage de la variable Q17
ewcs_2015_new$Q17 <- ewcs_2015_new$Q17
ewcs_2015_new$Q17[ewcs_2015_new$Q17 == "DK"] <- NA
ewcs_2015_new$Q17[ewcs_2015_new$Q17 == "Less than 1 year"] <- "0"
ewcs_2015_new$Q17[ewcs_2015_new$Q17 == "Not applicable (spontaneous)"] <- NA
ewcs_2015_new$Q17[ewcs_2015_new$Q17 == "Refusal"] <- NA
ewcs_2015_new$Q17 <- as.numeric(ewcs_2015_new$Q17)

## Recodage de la variable Q24
ewcs_2015_new$Q24 <- ewcs_2015_new$Q24
```

## Annexes A.

---

```
ewcs_2015_new$Q24[ewcs_2015_new$Q24 == "DK/no opinion"] <- NA
ewcs_2015_new$Q24 <- as.numeric(ewcs_2015_new$Q24)

## Recodage de la variable Q88
ewcs_2015_new$Q88 <- ewcs_2015_new$Q88
ewcs_2015_new$Q88[ewcs_2015_new$Q88 == "DK/no opinion (spontaneous)"] <- NA
ewcs_2015_new$Q88[ewcs_2015_new$Q88 == "Refusal (spontaneous)"] <- NA
ewcs_2015_new$Q88 <- factor(ewcs_2015_new$Q88)

## Recodage de la variable Q104_euro
ewcs_2015_new$Q104_euro <- ewcs_2015_new$Q104_euro
ewcs_2015_new$Q104_euro[ewcs_2015_new$Q104_euro == "DK (spontaneous)"] <- NA
ewcs_2015_new$Q104_euro[ewcs_2015_new$Q104_euro == "Refusal (spontaneous)"] <- NA
ewcs_2015_new$Q104_euro <- as.numeric(ewcs_2015_new$Q104_euro)

## Recodage de la variable ISCED
ewcs_2015_new$ISCED <- ewcs_2015_new$ISCED
ewcs_2015_new$ISCED[ewcs_2015_new$ISCED == "DK"] <- NA
ewcs_2015_new$ISCED[ewcs_2015_new$ISCED == "Refusal"] <- NA
ewcs_2015_new$ISCED <- factor(ewcs_2015_new$ISCED)

### Boxplot de la variable salaire afin de supprimé les valeurs extrême
boxplot(ewcs_2015_new$Q104_euro)
ewcs_2015_new <- na.omit(ewcs_2015_new)

# Renommé les variables pour mieux comprendre notre étude
ewcs_2015_ren <- ewcs_2015_new%>% rename(Pays = Country, genre = Q2a, age = Q2b, Taille_entrep =
                                             heure_par_semaine = Q24, secteur_trav = Q14, satisf =
                                             nivx_etude = ISCED, profession = isco_08_1)
summary(ewcs_2015_new$Q104_euro)

## remplacements par des valeurs manquantes les valeurs qui sont
## inferieur à 10% du quantile et superieur à 99% du quantile
## pour une bonne analyse de notre base de données
d <- quantile(ewcs_2015_ren$salaire, probs=seq(0, 1, 0.01),na.rm = T)
ewcs_2015_ren$salaire[ewcs_2015_ren$salaire <= d[["2%"]]] <-NA
ewcs_2015_ren$salaire[ewcs_2015_ren$salaire >= d[["99%"]]] <-NA

##### suppression des NA
ewcs_2015_ren <- na.omit(ewcs_2015_ren)

boxplot(ewcs_2015_ren$salaire)
summary(ewcs_2015_ren$salaire)

#####
```

## Annexes A.

---

```
## Rajout des colonnes (heures par mois et taux horaires)

*****Heure par mois ****#
heures_mois <- 4*ewcs_2015_ren$heure_par_semaine
names(heures_mois) <- "heures_mois"
ewcs_2015_ren <- cbind(ewcs_2015_ren,heures_mois)

## Taux horaires
*****Taux horaires*****
taux_horaires <- ewcs_2015_ren$salaire/ewcs_2015_ren$heures_mois
names(taux_horaires) <- "taux_horaires"
ewcs_2015_ren <- cbind(ewcs_2015_ren,taux_horaires)

## ordonner la base de données
## Réordonnement de ewcs_2015_ren$niveau_etude
ewcs_2015_ren$nivx_etude <- factor(ewcs_2015_ren$nivx_etude,
                                      levels = c("Doctorate or equivalent", "Master or equivalent",
                                                "Short-cycle tertiary education", "Post-secondary education",
                                                "Upper secondary education", "Lower secondary education",
                                                "Primary education"))

## Renommé le niveau d'étude
ewcs_2015_ren$nivx_etude <- as.character(ewcs_2015_ren$nivx_etude)
ewcs_2015_ren$nivx_etude[ewcs_2015_ren$nivx_etude == "Doctorate or equivalent"] <- "Doctorat"
ewcs_2015_ren$nivx_etude[ewcs_2015_ren$nivx_etude == "Master or equivalent"] <- "Master"
ewcs_2015_ren$nivx_etude[ewcs_2015_ren$nivx_etude == "Bachelor or equivalent"] <- "Licence"
ewcs_2015_ren$nivx_etude[ewcs_2015_ren$nivx_etude == "Short-cycle tertiary education"] <- "Enseignement supérieur"
ewcs_2015_ren$nivx_etude[ewcs_2015_ren$nivx_etude == "Post-secondary non-tertiary education"] <- "Enseignement secondaire"
ewcs_2015_ren$nivx_etude[ewcs_2015_ren$nivx_etude == "Upper secondary education"] <- "Lycée"
ewcs_2015_ren$nivx_etude[ewcs_2015_ren$nivx_etude == "Lower secondary education"] <- "Collège"
ewcs_2015_ren$nivx_etude[ewcs_2015_ren$nivx_etude == "Early childhood education"] <- "Maternité"
ewcs_2015_ren$nivx_etude[ewcs_2015_ren$nivx_etude == "Primary education"] <- "Primaire"
ewcs_2015_ren$nivx_etude <- factor(ewcs_2015_ren$nivx_etude)

## Recodage de la variable satisfaction
ewcs_2015_ren$satisfaction <- as.character(ewcs_2015_ren$satisfaction)
ewcs_2015_ren$satisfaction[ewcs_2015_ren$satisfaction == "Not at all satisfied"] <- "Pas du tout satisfait"
ewcs_2015_ren$satisfaction[ewcs_2015_ren$satisfaction == "Not very satisfied"] <- "Pas très satisfait"
ewcs_2015_ren$satisfaction[ewcs_2015_ren$satisfaction == "Satisfied"] <- "Satisfait"
ewcs_2015_ren$satisfaction[ewcs_2015_ren$satisfaction == "Very satisfied"] <- "Très satisfait"
ewcs_2015_ren$satisfaction <- factor(ewcs_2015_ren$satisfaction)

## Recodage de la variable type_contrat
ewcs_2015_ren$type_contrat <- as.character(ewcs_2015_ren$type_contrat)
ewcs_2015_ren$type_contrat[ewcs_2015_ren$type_contrat == ""] <- NA
ewcs_2015_ren$type_contrat[ewcs_2015_ren$type_contrat == "Full time"] <- "Temps plein"
ewcs_2015_ren$type_contrat[ewcs_2015_ren$type_contrat == "Part time"] <- "Temps partiel"
```

## Annexes A.

---

```
ewcs_2015_ren$type_contrat <- factor(ewcs_2015_ren$type_contrat)

#####
## Recodage de la variable genre
ewcs_2015_ren$genre <- as.character(ewcs_2015_ren$genre)
ewcs_2015_ren$genre[ewcs_2015_ren$genre == "Female"] <- "Femme"
ewcs_2015_ren$genre[ewcs_2015_ren$genre == "Male"] <- "Homme"
ewcs_2015_ren$genre <- factor(ewcs_2015_ren$genre)

## Recodage de la variable situation
ewcs_2015_ren$situation <- as.character(ewcs_2015_ren$situation)
ewcs_2015_ren$situation[ewcs_2015_ren$situation == "Employee"] <- "Employé"
ewcs_2015_ren$situation[ewcs_2015_ren$situation == "Self-employed"] <- "Travailleur_Indép"
ewcs_2015_ren$situation <- factor(ewcs_2015_ren$situation)

## Decoupage des tranches d'age par quantile
ewcs_2015_ren$age_rec <- cut(ewcs_2015_ren$age,
                           include.lowest = FALSE,
                           right = FALSE,
                           dig.lab = 4,
                           breaks = c(18, 27, 37, 46, 56, 65, 75, 100))

## Recodage de la variable annee_experience en intervalle qui sera notée annee_experience_re
ewcs_2015_ren$ancienneté_rec <- cut(ewcs_2015_ren$ancienneté,
                                      include.lowest = FALSE,
                                      right = FALSE,
                                      dig.lab = 4,
                                      breaks = c(0, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50))

## Recodage de la variable secteur
ewcs_2015_ren$secteur_trav <- as.character(ewcs_2015_ren$secteur_trav)
ewcs_2015_ren$secteur_trav[ewcs_2015_ren$secteur_trav == "A joint private-public organisation"] <- "Autre"
ewcs_2015_ren$secteur_trav[ewcs_2015_ren$secteur_trav == "Other, please specify:" ] <- "Autre"
ewcs_2015_ren$secteur_trav[ewcs_2015_ren$secteur_trav == "The not-for-profit sector or an NGO"] <- "sect_nfp"
ewcs_2015_ren$secteur_trav[ewcs_2015_ren$secteur_trav == "The private sector"] <- "sect_priv"
ewcs_2015_ren$secteur_trav[ewcs_2015_ren$secteur_trav == "The public sector"] <- "sect_publi"
ewcs_2015_ren$secteur_trav <- factor(ewcs_2015_ren$secteur_trav)

## Recodage de la variable Profession
ewcs_2015_ren$profession <- as.character(ewcs_2015_ren$profession)
ewcs_2015_ren$profession[ewcs_2015_ren$profession == ""] <- NA
ewcs_2015_ren$profession <- factor(ewcs_2015_ren$profession)

## Recodage de ewcs_2015_fra$profession
```

## Annexes A.

---

```
ewcs_2015_ren$profession <- as.character(ewcs_2015_ren$profession)
ewcs_2015_ren$profession[ewcs_2015_ren$profession == "Armed forces occupations"] <- "Professionnelles"
ewcs_2015_ren$profession[ewcs_2015_ren$profession == "Clerical support workers"] <- "Employés"
ewcs_2015_ren$profession[ewcs_2015_ren$profession == "Craft and related trades workers"] <- "Professions"
ewcs_2015_ren$profession[ewcs_2015_ren$profession == "Elementary occupations"] <- "Professions"
ewcs_2015_ren$profession[ewcs_2015_ren$profession == "Plant and machine operators, and assemblers"] <- "Professionnels"
ewcs_2015_ren$profession[ewcs_2015_ren$profession == "Professionals"] <- "Professionnels"
ewcs_2015_ren$profession[ewcs_2015_ren$profession == "Service and sales workers"] <- "Personnel de service et de vente"
ewcs_2015_ren$profession[ewcs_2015_ren$profession == "Skilled agricultural, forestry and fishery workers"] <- "Professionnels"
ewcs_2015_ren$profession[ewcs_2015_ren$profession == "Technicians and associate professionals"] <- "Professionnels"
ewcs_2015_ren$profession <- factor(ewcs_2015_ren$profession)

## Suppression de toutes les valeurs NA
ewcs_2015_ren <- na.omit(ewcs_2015_ren)

#irec(ewcs_2015_ren, "profession")

## Recodage de ewcs_2015_ren$profession en ewcs_2015_ren$profession_rec
ewcs_2015_ren$profession_rec <- as.character(ewcs_2015_ren$profession)
ewcs_2015_ren$profession_rec[ewcs_2015_ren$profession == "Artisans_et_travailleurs_des_métiers"] <- "A"
ewcs_2015_ren$profession_rec[ewcs_2015_ren$profession == "Conducteurs_d_installations_et_de_machines"] <- "B"
ewcs_2015_ren$profession_rec[ewcs_2015_ren$profession == "Employés_de_soutien_administratif"] <- "C"
ewcs_2015_ren$profession_rec[ewcs_2015_ren$profession == "Managers"] <- "D"
ewcs_2015_ren$profession_rec[ewcs_2015_ren$profession == "Personnel_de_service_et_de_vente"] <- "E"
ewcs_2015_ren$profession_rec[ewcs_2015_ren$profession == "Professionnels"] <- "F"
ewcs_2015_ren$profession_rec[ewcs_2015_ren$profession == "Professions_des_forces_armées"] <- "G"
ewcs_2015_ren$profession_rec[ewcs_2015_ren$profession == "Professions_elémentaires"] <- "H"
ewcs_2015_ren$profession_rec[ewcs_2015_ren$profession == "Techniciens_et_professionnels_associés"] <- "I"
ewcs_2015_ren$profession_rec <- factor(ewcs_2015_ren$profession_rec)

##### fin du pretraitement #####
#####

#####
##### Choix des deux pays #####
#####
### France premier pays
ewcs_2015_fra <- subset(ewcs_2015_ren, Pays == "France")

## L'Allemagne deuxième pays
ewcs_2015_ger <- subset(ewcs_2015_ren, Pays == "Germany")
```

## Annexes A.

---

```
#####
##### Description de la base des deux pays #####
#####

##### France #####
## Affichage d'une partie de la base
head(ewcs_2015_fra)
dim(ewcs_2015_fra) ## Taille de la base

##### Summary de la base de donnée
summary(ewcs_2015_fra)

##### Statistique de la base de données #####
## Pour La france:

## Statt du salaire
summary(ewcs_2015_fra$salaire)

## salaire en fonction du genre
by(ewcs_2015_fra[, c("salaire")], INDICES=ewcs_2015_fra[,c("genre")], FUN = summary, na.rm = T)

## ecart type
by(ewcs_2015_fra[, c("salaire")], INDICES=ewcs_2015_fra[,c("genre")], FUN = sd, na.rm = T)

## salaire en fonction du niveau d'étude
by(ewcs_2015_fra[, c("salaire")], INDICES=ewcs_2015_fra[,c("genre", "nivx_etude")], FUN = sum)

## ecart type
by(ewcs_2015_fra[, c("salaire")], INDICES=ewcs_2015_fra[,c("genre", "nivx_etude")], FUN = sd, na.rm = T)

## salaire en fonction de la taille de l'entreprise
by(ewcs_2015_fra[, c("salaire")], INDICES=ewcs_2015_fra[,c("genre", "nivx_etude")], FUN = sum)

## ecart type
by(ewcs_2015_fra[, c("salaire")], INDICES=ewcs_2015_fra[,c("genre", "nivx_etude")], FUN = sd, na.rm = T)

## salaire en fonction de la profession
by(ewcs_2015_fra[, c("salaire")], INDICES=ewcs_2015_fra[,c("genre", "profession")], FUN = sum)

## ecart type
by(ewcs_2015_fra[, c("salaire")], INDICES=ewcs_2015_fra[,c("genre", "profession")], FUN = sd, na.rm = T)

## salaire en fonction du type de contrat
by(ewcs_2015_fra[, c("salaire")], INDICES=ewcs_2015_fra[,c("genre", "type_contrat")], FUN = sum)

## salaire en fonction du type de contrat
by(ewcs_2015_fra[, c("salaire")], INDICES=ewcs_2015_fra[,c("type_contrat")], FUN = summary,
```

## Annexes A.

---

```
## ecart type
by(ewcs_2015_fra[, c("salaire")], INDICES=ewcs_2015_fra[,c("genre","type_contrat")], FUN = s

## salaire en fonction de la situation
by(ewcs_2015_fra[, c("salaire")], INDICES=ewcs_2015_fra[,c("genre", "situation")], FUN = sum)

## salaire en fonction de la situation
by(ewcs_2015_fra[, c("salaire")], INDICES=ewcs_2015_fra[,c("situation")], FUN = summary, na.rm = T)

## ecart type
by(ewcs_2015_fra[, c("salaire")], INDICES=ewcs_2015_fra[,c("genre","situation")], FUN = sd, na.rm = T

## salaire en fonction du secteur
by(ewcs_2015_fra[, c("salaire")], INDICES=ewcs_2015_fra[,c("genre", "secteur_trav")], FUN = sum)

## ecart type
by(ewcs_2015_fra[, c("salaire")], INDICES=ewcs_2015_fra[,c("genre","secteur_trav")], FUN = sd, na.rm = T

## salaire en fonction du secteur
by(ewcs_2015_fra[, c("salaire")], INDICES=ewcs_2015_fra[,c("genre", "Taille_entreprise")], FUN = sum)

## ecart type
by(ewcs_2015_fra[, c("salaire")], INDICES=ewcs_2015_fra[,c("genre","Taille_entreprise")], FUN = sd, na.rm = T

##### Allemagne #####
## Affichage d'une partie de la base
head(ewcs_2015_ger)

## Affichage de la dimension de la base de données
dim(ewcs_2015_ger)

##### Summary de la base de donnée
summary(ewcs_2015_ger)

## Statt du salaire
summary(ewcs_2015_ger$salaire)

## salaire en fonction du genre
by(ewcs_2015_ger[, c("salaire")], INDICES=ewcs_2015_ger[,c("genre")], FUN = summary, na.rm = T)

## ecart type
by(ewcs_2015_ger[, c("salaire")], INDICES=ewcs_2015_ger[,c("genre")], FUN = sd, na.rm = T

## salaire en fonction du niveau d'étude
by(ewcs_2015_ger[, c("salaire")], INDICES=ewcs_2015_ger[,c("genre","nivx_etude")], FUN = sum)

## ecart type
```

## Annexes A.

---

```
by(ewcs_2015_ger[, c("salaire")], INDICES=ewcs_2015_ger[,c("genre","nivx_etude")], FUN = sd,  
## salaire en fonction du type de contrat  
by(ewcs_2015_ger[, c("salaire")], INDICES=ewcs_2015_ger[,c("genre","type_contrat")], FUN = s  
## salaire en fonction du type de contrat  
by(ewcs_2015_ger[, c("salaire")], INDICES=ewcs_2015_ger[,c("type_contrat")], FUN = summary,  
  
## ecart type  
by(ewcs_2015_ger[, c("salaire")], INDICES=ewcs_2015_ger[,c("genre","type_contrat")], FUN = s  
## salaire en fonction de la profession  
by(ewcs_2015_ger[, c("salaire")], INDICES=ewcs_2015_ger[,c("genre","profession")], FUN = sum  
## ecart type  
by(ewcs_2015_ger[, c("salaire")], INDICES=ewcs_2015_ger[,c("genre","profession")], FUN = sd,  
## l'Allemagne salaire en fonction de la taille de l'entreprise  
by(ewcs_2015_ger[, c("salaire")], INDICES=ewcs_2015_ger[,c("genre","secteur_trav")], FUN = s  
## ecart type  
by(ewcs_2015_ger[, c("salaire")], INDICES=ewcs_2015_ger[,c("genre","secteur_trav")], FUN = s  
## salaire en fonction du genre  
by(ewcs_2015_ger[, c("salaire")], INDICES=ewcs_2015_ger[,c("genre", "situation")], FUN = sum  
## salaire en fonction du genre  
by(ewcs_2015_ger[, c("salaire")], INDICES=ewcs_2015_ger[,c("situation")], FUN = summary, na.rm=TRUE  
## ecart type  
by(ewcs_2015_ger[, c("salaire")], INDICES=ewcs_2015_ger[,c("genre", "situation")], FUN = sd,  
  
## l'Allemagne salaire en fonction de la taille de l'entreprise  
by(ewcs_2015_ger[, c("salaire")], INDICES=ewcs_2015_ger[,c("genre","Taille_entreprise")], FUN = s  
## ecart type  
by(ewcs_2015_ger[, c("salaire")], INDICES=ewcs_2015_ger[,c("genre","Taille_entreprise")], FUN = s  
  
## l'Allemagne salaire en fonction de la Profession  
by(ewcs_2015_ger[, c("salaire")], INDICES=ewcs_2015_ger[,c("genre","profession_rec")], FUN = s  
## ecart type  
by(ewcs_2015_ger[, c("salaire")], INDICES=ewcs_2015_ger[,c("genre","profession_rec")], FUN = s  
##### proportion des individus en fonction d'un certain nombres de variables#####  
#### proportion des individus en fonction d'un certain nombres de variables#####
```

## Annexes A.

---

```
#### Proportion des individus #####
#####
## Repartition des individus
table(ewcs_2015_fra$genre)
table(ewcs_2015_ger$genre)

## Proportion en % des individus en fonction du salaire
table(ewcs_2015_fra$genre,ewcs_2015_fra$salaire)
round(100*prop.table(table(ewcs_2015_fra$genre,ewcs_2015_fra$salaire), margin = 1),2)

## Proportion en % des individus en fonction du salaire
table(ewcs_2015_ger$genre,ewcs_2015_ger$salaire)
round(100*prop.table(table(ewcs_2015_ger$genre,ewcs_2015_ger$salaire), margin = 1),2)

## Proportion en % des individus en fonction du type de contrat
table(ewcs_2015_fra$genre,ewcs_2015_fra$type_contrat)
round(100*prop.table(table(ewcs_2015_fra$genre,ewcs_2015_fra$type_contrat), margin = 1),2)

## Proportion en % des individus en fonction du type de contrat
table(ewcs_2015_fra$genre,ewcs_2015_fra$type_contrat,ewcs_2015_fra$profession)
round(100*prop.table(table(ewcs_2015_fra$genre,ewcs_2015_fra$type_contrat,ewcs_2015_fra$prof

## Proportion en % des individus en fonction du type de contrat
table(ewcs_2015_ger$genre,ewcs_2015_ger$type_contrat)
round(100*prop.table(table(ewcs_2015_ger$genre,ewcs_2015_ger$type_contrat), margin = 1),2)

### Proportion en % des individus en fonction du privée et du public
table(ewcs_2015_fra$genre,ewcs_2015_fra$secteur_trav)
round(100*prop.table(table(ewcs_2015_fra$genre,ewcs_2015_fra$secteur_trav), margin = 1),2)

### Proportion en % des individus en fonction du privée et du public
table(ewcs_2015_ger$genre,ewcs_2015_ger$secteur_trav)
round(100*prop.table(table(ewcs_2015_ger$genre,ewcs_2015_ger$secteur_trav), margin = 1),2)

## Proposition en % des individu en fonction des ages
table(ewcs_2015_fra$genre,ewcs_2015_fra$age_rec)
round(100*prop.table(table(ewcs_2015_fra$genre,ewcs_2015_fra$age_rec), margin = 1),2)

## Proposition en % des individu en fonction des ages
table(ewcs_2015_ger$genre,ewcs_2015_ger$age_rec)
round(100*prop.table(table(ewcs_2015_ger$genre,ewcs_2015_ger$age_rec), margin = 1),2)

## Proposition des individus en fonction de la profession
table(ewcs_2015_fra$genre,ewcs_2015_fra$profession_rec)
```

## Annexes A.

---

```
round(100*prop.table(table(ewcs_2015_fra$genre,ewcs_2015_fra$profession_rec),margin = 1), 2)

## Proposition des individus en fonction de la profession
table(ewcs_2015_ger$genre,ewcs_2015_ger$profession_rec)
round(100*prop.table(table(ewcs_2015_ger$genre,ewcs_2015_ger$profession_rec),margin = 1), 2)

## Proportion des individus en fonction du niveau d'étude
table(ewcs_2015_fra$genre,ewcs_2015_fra$nivx_etude)
round(100*prop.table(table(ewcs_2015_fra$genre,ewcs_2015_fra$nivx_etude),margin = 1), 2)

## Proposition des individus en fonction de la profession et du niveau d'étude
table(ewcs_2015_ger$genre,ewcs_2015_ger$nivx_etude)
round(100*prop.table(table(ewcs_2015_ger$genre,ewcs_2015_ger$nivx_etude),margin = 1), 2)

## Proportion des individu en fonction de la situation
tab4 = table(ewcs_2015_ger$genre, ewcs_2015_ger$situation)
round(100*prop.table(tab4, margin = 1), 2)

## Proportion des individu en fonction de la situation
tab4 = table(ewcs_2015_fra$genre, ewcs_2015_fra$situation)
round(100*prop.table(tab4, margin = 1), 2)

##### Proportion des individu en fonction de la satisfaction
t5 = table(ewcs_2015_fra$genre,ewcs_2015_fra$satisfaction)
round(100*prop.table(t5,margin=1),2)

##### Proportion des individu en fonction de la satisfaction
t5 = table(ewcs_2015_ger$genre,ewcs_2015_ger$satisfaction)
round(100*prop.table(t5,margin=1),2)

## Proposition des individu en fonction de l'ancienneté
t4=table(ewcs_2015_fra$genre,ewcs_2015_fra$ancienneté_rec)
round(100*prop.table(t4,margin=1),2)

## Proposition des individu en fonction de l'ancienneté
t4=table(ewcs_2015_ger$genre,ewcs_2015_ger$ancienneté_rec)
round(100*prop.table(t4,margin=1),2)

## Proposition des individu en fonction de la profession
t4=table(ewcs_2015_fra$genre,ewcs_2015_fra$profession_rec)
round(100*prop.table(t4,margin=1),2)

## Proposition des individu en fonction de la Profession
t4=table(ewcs_2015_ger$genre,ewcs_2015_ger$profession_rec)
round(100*prop.table(t4,margin=1),2)
```

## Annexes A.

---

```
#####
#####***** Datavisualisation ****#
#####

coul <- brewer.pal(8, "Set2")

##### Visualisation de la repartion du genre en fonction des variables

### Repartion des individus en genre (France)
barplot(table(ewcs_2015_fra$genre),col = coul, main = paste("Repartition des", nrow(ewcs_2015_fra),
xlab = "genre", ylab = "Nombres d'individus" )

### Repartion des individus en genre (Allemagne)
barplot(table(ewcs_2015_ger$genre),col = coul, main = paste("Repartition des", nrow(ewcs_2015_ger),
xlab = "genre", ylab = "Nombres d'individus" )

ggplot(data = ewcs_2015_ger)+aes(x = genre)+geom_bar(aes(fill = genre))+labs(ylab = "Nombres d'individus", title = "Repartition des individus en fonction du genre", fill = "Genre")

X11()
## Histogramme du salaire (France)
hist(ewcs_2015_fra$salaire,col = coul, main = "Répartition du salaire (France)",
xlab = "Salaires en euro",las=2,labels = TRUE)

## Histogramme du salaire (Allemagne)
hist(ewcs_2015_ger$salaire,las=2,col = coul, main = "Repartition du salaire (Allemagne)",
xlab = "Salaires en euro",labels = TRUE )

##### histogramme de la repartition des individus en fonction des ages (France)
ggplot(data = ewcs_2015_fra)+aes(x = age_rec)+geom_bar(aes(fill = genre))+ggtile("Repartition des ages en fonction du nombre d'individu")

##### histogramme de la repartition des individus en fonction des ages (Allemagne)
ggplot(data = ewcs_2015_ger)+aes(x = age_rec)+geom_bar(aes(fill = age_rec))+ggtile("Repartition des ages en fonction du nombre d'individu")

## Repartition des individus en fonction de la satisfaction (France)
plot(table(ewcs_2015_fra$satisfaction), col=coul,las=1 ,
main = paste("Répartition des", nrow(ewcs_2015_fra), "individus en fonction de la satisfaction"))

## Repartition des individus en fonction de la satisfaction (Allemagne)
plot(table(ewcs_2015_ger$satisfaction), col=coul,las=1 ,
main = paste("Répartition des", nrow(ewcs_2015_ger), "individus en fonction de la satisfaction"))

## Repartition des individus en fonction du niveau d'étude (France)
barplot(height=table(ewcs_2015_fra$nivx_etude), col=coul,las=2 ,
main = paste("Répartition des", nrow(ewcs_2015_fra), "individus en fonction du diplôme"))

## Repartition des individus en fonction du niveau d'étude (Allemagne)
barplot(height=table(ewcs_2015_ger$nivx_etude), col=coul,las=2 ,
```

## Annexes A.

---

```
main = paste("Répartition des", nrow(ewcs_2015_ger), "individus en fonction du diplome", sep="")  
  
## Repartition des individus en fonction du type de contrat (France)  
barplot(height=table(ewcs_2015_fra$type_contrat), col=coul,las=1 ,  
        main = paste("Répartition des", nrow(ewcs_2015_fra), "individus en fonction du type de contrat", sep="")  
  
## Repartition des individus en fonction du type de contrat (Allemagne)  
barplot(table(ewcs_2015_ger$type_contrat), col=coul,las=1,  
        main = paste("Répartition des", nrow(ewcs_2015_ger), "individus en fonction du type de contrat", sep="")  
  
## Repartition des individus en fonction de la situation (France)  
barplot(height=table(ewcs_2015_fra$situation), col=coul,las=1 ,  
        main = paste("Répartition des", nrow(ewcs_2015_fra), "individus en fonction de leur situation", sep="")  
  
## Repartition des individus en fonction de la situation (Allemagne)  
barplot(table(ewcs_2015_ger$situation), col=coul,las=1,  
        main = paste("Répartition des", nrow(ewcs_2015_ger), "individus en fonction de leur situation", sep="")  
  
## Repartition des individus en fonction du secteur (France)  
barplot(height=table(ewcs_2015_fra$secteur_trav), col=coul,las=1 ,  
        main = paste("Répartition des", nrow(ewcs_2015_fra), "individus en fonction du secteur", sep="")  
  
## Repartition des individus en fonction du secteur (Allemagne)  
barplot(table(ewcs_2015_ger$secteur_trav), col=coul,las=1 ,  
        main = paste("Répartition des", nrow(ewcs_2015_ger), "individus en fonction du secteur", sep="")  
  
## Repartition des individus en fonction de l'ancienneté (France)  
barplot(height=table(ewcs_2015_fra$ancienneté_rec), col=coul,las=1 ,  
        main = paste("Répartition des", nrow(ewcs_2015_fra), "individus en fonction de l'ancienneté", sep="")  
  
## Repartition des individus en fonction de l'ancienneté (Allemagne)  
barplot(table(ewcs_2015_ger$ancienneté_rec), col=coul,las=1 ,  
        main = paste("Répartition des", nrow(ewcs_2015_ger), "individus en fonction de l'ancienneté", sep="")  
  
## Repartition des individus en fonction de l'age (France)  
barplot(height=table(ewcs_2015_fra$age_rec), col=coul,las=1 ,  
        main = paste("Répartition des", nrow(ewcs_2015_fra), "individus en fonction de l'age", sep="")  
  
## Repartition des individus en fonction de l'age (Allemagne)  
barplot(table(ewcs_2015_ger$age_rec), col=coul,las=1 ,  
        main = paste("Répartition des", nrow(ewcs_2015_ger), "individus en fonction de l'age", sep="")  
  
#####  
### Graphique en mosaïplot pour voir la repartition des du genre en fonction du niveau d'étude  
mosaicplot(nivx_etude ~ genre, data = ewcs_2015_fra, col=c("green","cyan"),xlab="Niveaux d'éducation")  
  
### Graphique en mosaïplot pour voir la repartition des du genre en fonction du niveau d'étude  
mosaicplot(nivx_etude ~ genre, data = ewcs_2015_ger, col=c("green","cyan"),xlab="Niveaux d'éducation")
```

## Annexes A.

---

```
### Graphique en mosaïplot pour voir la répartition des genres en fonction du type de contrat
mosaicplot(type_contrat ~ genre, data = ewcs_2015_fra, col=c("green","cyan")),xlab="Genres",ylab="Genre"

### Graphique en mosaïplot pour voir la répartition des genres en fonction du type de contrat
mosaicplot(type_contrat ~ genre, data = ewcs_2015_ger, col=c("green","cyan")),xlab="Genres",ylab="Genre"

### Graphique en mosaïplot pour voir la répartition des genres en fonction du secteur(France)
mosaicplot(secteur_trav ~ genre, data = ewcs_2015_fra, col=c("green","cyan")),xlab="Genres",ylab="Genre"

### Graphique en mosaïplot pour voir la répartition des genres en fonction du secteur(Allemagne)
mosaicplot(secteur_trav ~ genre, data = ewcs_2015_ger, col=c("green","cyan")),xlab="Genres",ylab="Genre"

### Graphique en mosaïplot pour voir la répartition des genres en fonction de la satisfaction
mosaicplot(profession_rec~satisfaction , data = ewcs_2015_fra, col=c("green","cyan"), "yellow", "grey"))

### Graphique en mosaïplot pour voir la répartition des genres en fonction de la satisfaction
mosaicplot(profession_rec~satisfaction, data = ewcs_2015_ger,col=c("green","cyan", "yellow", "grey"))

### Graphique en mosaïplot pour voir la répartition des genres en fonction de la situation(France)
mosaicplot(situation ~ genre, data = ewcs_2015_fra, col=c("green","cyan")),xlab="Genres",ylab="Genre"

### Graphique en mosaïplot pour voir la répartition des genres en fonction de la situation(Angleterre)
mosaicplot(situation ~ genre, data = ewcs_2015_ger, col=c("green","cyan")),xlab="Genres",ylab="Genre"

### Graphique en mosaïplot pour voir la répartition des genres en fonction de la profession(France)
mosaicplot(profession_rec ~ genre, data = ewcs_2015_fra, col=c("green","cyan")),xlab="Profession",ylab="Genre"

### Graphique en mosaïplot pour voir la répartition des genres en fonction de la profession(Angleterre)
mosaicplot(profession_rec ~ genre, data = ewcs_2015_ger, col=c("green","cyan")),xlab="Profession",ylab="Genre"

## Visualisation des salaires en fonction du genre
ggplot(ewcs_2015_fra, aes(x = genre, y = salaire, fill = genre)) +
  geom_boxplot(outlier.colour="Gold", outlier.shape=16, outlier.size=2)+ggtitle("Répartition des salaires par genre")

## (ALL) Visualisation des salaires en fonction du genre
ggplot(ewcs_2015_ger, aes(x = genre, y = salaire, fill = genre)) +
  geom_boxplot(outlier.colour="Gold", outlier.shape=16, outlier.size=2)+ggtitle("Répartition des salaires par genre")

## Visualisation des salaires en fonction de la situation et du genre
ggplot(ewcs_2015_fra, aes(x = situation, y = salaire, fill = genre)) +
  geom_boxplot(outlier.colour="Gold", outlier.shape=16, outlier.size=2)+ggtitle("Répartition des salaires par situation et genre")

## (ALL) Visualisation des salaires en fonction de la situation et du genre
ggplot(ewcs_2015_ger, aes(x = situation, y = salaire, fill = genre)) +
  geom_boxplot(outlier.colour="Gold", outlier.shape=16, outlier.size=2)+ggtitle("Répartition des salaires par situation et genre")
```

```
## Visualisation des salaire en fonction du contrat
ggplot(ewcs_2015_fra, aes(x = type_contrat, y = salaire, fill = genre)) +
  geom_boxplot(outlier.colour="Gold", outlier.shape=16, outlier.size=2)+ggtitle("Repartition des salaires par type de contrat en France")

## (ALL) Visualisation des salaire en fonction du contrat
ggplot(ewcs_2015_ger, aes(x = type_contrat, y = salaire, fill = genre)) +
  geom_boxplot(outlier.colour="Gold", outlier.shape=16, outlier.size=2)+ggtitle("Repartition des salaires par type de contrat en Allemagne")

## Visualisation des salaire en fonction du niveau d'études
ggplot(ewcs_2015_fra, aes(x = nivx_etude, y = salaire, fill = genre)) +
  geom_boxplot(outlier.colour="Green", outlier.shape=16, outlier.size=2)+ggtitle("Repartition des salaires par niveau d'études en France")

## (ALL) Visualisation des salaire en fonction du contrat
ggplot(ewcs_2015_ger, aes(x = nivx_etude, y = salaire, fill = genre)) +
  geom_boxplot(outlier.colour="Gold", outlier.shape=16, outlier.size=2)+ggtitle("Repartition des salaires par niveau d'études en Allemagne")

## Visualisation des salaire en fonction du contrat
ggplot(ewcs_2015_fra, aes(x = Taille_entreprise, y = salaire, fill = genre)) +
  geom_boxplot(outlier.colour="Green", outlier.shape=16, outlier.size=2)+ggtitle("Repartition des salaires par taille d'entreprise en France")

## (ALL) Visualisation des salaire en fonction du contrat
ggplot(ewcs_2015_ger, aes(x = Taille_entreprise, y = salaire, fill = genre)) +
  geom_boxplot(outlier.colour="Gold", outlier.shape=16, outlier.size=2)+ggtitle("Repartition des salaires par taille d'entreprise en Allemagne")

## Visualisation des salaire en fonction de l'ancienneté et le genre
ggplot(ewcs_2015_fra, aes(x = secteur_trav, y = salaire, fill = genre)) +
  geom_boxplot(outlier.colour="Green", outlier.shape=16, outlier.size=2)+ggtitle("Repartition des salaires par secteur d'activité en France")

## (ALL) Visualisation des salaire en fonction de l'ancienneté et le genre
ggplot(ewcs_2015_ger, aes(x = secteur_trav, y = salaire, fill = genre)) +
  geom_boxplot(outlier.colour="Gold", outlier.shape=16, outlier.size=2)+ggtitle("Repartition des salaires par secteur d'activité en Allemagne")

## Visualisation des salaire en fonction de la satisfaction et le genre
ggplot(ewcs_2015_fra, aes(x = satisfaction, y = salaire, fill = genre)) +
  geom_boxplot(outlier.colour="Green", outlier.shape=16, outlier.size=2)+ggtitle("Repartition des salaires par satisfaction en France")

## (ALL) Visualisation des salaire en fonction de la satisfaction et le genre
ggplot(ewcs_2015_ger, aes(x = satisfaction, y = salaire, fill = genre)) +
  geom_boxplot(outlier.colour="Gold", outlier.shape=16, outlier.size=2)+ggtitle("Repartition des salaires par satisfaction en Allemagne")
```

## Annexes A.

---

```
## Visualisation des salaire en fonction de la satisfaction et le genre
ggplot(ewcs_2015_fra, aes(x = ancienneté_rec, y = salaire, fill = genre)) +
  geom_boxplot(outlier.colour="blue", outlier.shape=16, outlier.size=2)+ggtitle("Repartition des salaires par ancienneté et genre en France")

## (ALL) Visualisation des salaire en fonction de la satisfaction et le genre
ggplot(ewcs_2015_ger, aes(x = ancienneté_rec, y = salaire, fill = genre)) +
  geom_boxplot(outlier.colour="Gold", outlier.shape=16, outlier.size=2)+ggtitle("Repartition des salaires par ancienneté et genre en Allemagne")

#####
##### Corrélation #####
#####

##### France #####
coul <- brewer.pal(8, "Set2")
### La France
m = ewcs_2015_fra[,c("age", "salaire", "ancienneté","heure_par_semaine",
                      "heures_mois", "taux_horaires")]
corr <- round(cor(m),1)
ggcorrplot(corr, p.mat = cor_pmat(ewcs_2015_fra[,c("age", "salaire", "ancienneté","heure_par_semaine",
                                                    "lower", color=c("#00AFBB","cyan","#FC4E07"), outline.col = "black", lab = TRUE)

##### Allemagne #####
m1 = ewcs_2015_ger[,c("age", "salaire", "ancienneté","heure_par_semaine",
                      "heures_mois", "taux_horaires")]
corr <- round(cor(m1),1)
ggcorrplot(corr, p.mat = cor_pmat(ewcs_2015_ger[,c("age", "salaire", "ancienneté","heure_par_semaine",
                                                    "lower", color=c("#00AFBB","cyan","#FC4E07"), outline.col = "black", lab = TRUE

#####
##### Analyse des données AFMD #####
#####

##### France #####
coul <- brewer.pal(8, "Set2")

data = ewcs_2015_fra[,c("genre","satisfaction","age","nivx_etude","profession_rec","situation",
                        "type_contrat", "secteur_trav", "Taille_entreprise",
                        "ancienneté","heure_par_semaine","heures_mois",
                        "salaire","taux_horaires")]

res.famd <- FAMD(data,ncp = 14,graph = FALSE)
round(head(get_eigenvalue(res.famd)),2)
```

```
## Visualiser les proportions de variances expliquées par les différents axes
fviz_screepplot(res.famd,main="Visualisation des proportions de variances en fonction des di
                                ylab="Pourcentage des variances expliquées",addlabels=TRUE, hjust = -0.3)

# Graphique des variables
X11()
fviz_famd_var (res.famd, repel = TRUE)

# Contribution à la première dimension
fviz_contrib (res.famd, "var", axes = 1)
# Contribution à la deuxième dimension
fviz_contrib (res.famd, "var", axes = 2)

#Pour extraire les résultats pour les variables quantitatives
quanti.var <- get_famd_var(res.famd, "quanti.var")

round(head(quanti.var$coord),2)
round(head(quanti.var$contrib),2)
round(head(quanti.var$cos2),2)
#round(head(quanti.var$cor),2)

fviz_famd_var(res.famd, "quanti.var", col.var = "contrib",
              gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
              repel = TRUE)

# Couleur par valeurs cos2: qualité sur le plan des facteurs
X11()
fviz_famd_var(res.famd, "quanti.var", col.var = "cos2",
              gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
              repel = TRUE)

#Comme les variables quantitatives, les résultats pour les variables qualitatives:
quali.var <- get_famd_var(res.famd, "quali.var")

head(quali.var$coord)
head(quali.var$contrib)
head(quali.var$cos2)
head(quali.var$cor)

# Contribution à la première dimension
fviz_contrib (res.famd, "quali.var", axes = 1)
# Contribution à la deuxième dimension
fviz_contrib (res.famd, "quali.var", axes = 2)
#Pour visualiser les variables qualitatives
fviz_famd_var(res.famd, "quali.var", col.var = "contrib",
              gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"))
```

## Annexes A.

---

```
##Graphique des individus
## Pour obtenir les résultats pour les individus:
ind <- get_famd_ind(res.famd)

#Pour visualiser les individus,
fviz_famd_ind(res.famd, col.ind = "cos2",
               gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
               repel = TRUE)

## Notez qu'il est possible de colorer les individus en utilisant l'une des variables qualitatives
fviz_mfa_ind(res.famd,
              habillage = "genre", # color by groups
              palette = c("#00AFBB", "#E7B800", "#FC4E07"),
              addEllipses = TRUE, ellipse.type = "confidence",
              repel = TRUE )

fviz_ellipses(res.famd, c("genre", "type_contrat"), repel = TRUE)

#####
#### Allemagne #####
#####

df = ewcs_2015_ger[,c("genre", "nivx_etude", "situation", "satisfaction", "profession_rec", "type",
                      "secteur_trav", "Taille_entreprise",
                      "age", "ancienneté", "heure_par_semaine", "heures_mois",
                      "salaire", "taux_horaires")]

res.famd <- FAMD(df, ncp = 14, graph = FALSE)

## La proportion de variances expliquées par les différentes dimensions (axes)
eig.val <- get_eigenvalue(res.famd)
round(head(eig.val), 2)

## Visualiser les proportions de variances expliquées par les différents axes
fviz_screeplot(res.famd, main="Visualisation des proportions de variances en fonction des dimensions",
               addlabels=TRUE, hjust = -0.3)

## taper la commande suivante pour éviter les erreurs
x11()
# Graphique des variables
fviz_famd_var (res.famd, repel = TRUE)

# Contribution à la première dimension
fviz_contrib (res.famd, "var", axes = 1)
# Contribution à la deuxième dimension
fviz_contrib (res.famd, "var", axes = 2)

#Pour extraire les résultats pour les variables quantitatives
```

```
quanti.var <- get_famd_var(res.famd, "quanti.var")
round(head(quanti.var$coord),2)
round(head(quanti.var$contrib),2)
round(head(quanti.var$cos2),2)
round(head(quanti.var$cor),2)

# Pour visualiser les variables quantitatives
fviz_famd_var(res.famd, "quanti.var", col.var = "contrib",
               gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
               repel = TRUE)

# Couleur par valeurs cos2: qualité sur le plan des facteurs
fviz_famd_var(res.famd, "quanti.var", col.var = "cos2",
               gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
               repel = TRUE)

#Comme les variables quantitatives, les résultats pour les variables qualitatives:
quali.var <- get_famd_var(res.famd, "quali.var")

round(head(quali.var$coord),2)
round(head(quali.var$contrib),2)
round(head(quali.var$cos2),2)
round(head(quali.var$cor),2)

#Pour visualiser les variables qualitatives
fviz_famd_var(res.famd, "quali.var", col.var = "contrib",
               gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"))

##Graphique des individus
## Pour obtenir les résultats pour les individus:
ind <- get_famd_ind(res.famd)

#Pour visualiser les individus,
X11()
fviz_famd_ind(res.famd, col.ind = "cos2",
               gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
               repel = TRUE)

## Notez qu'il est possible de colorer les individus en utilisant l'une des variables qualitatives
#X11()
fviz_mfa_ind(res.famd,
              habillage = "genre", # color by groups
              palette = c("#00AFBB", "#E7B800", "#FC4E07"),
              addEllipses = TRUE, ellipse.type = "confidence",
              repel = TRUE )

fviz_ellipses(res.famd, c("genre", "nivx_etude"), repel = TRUE)
```

## Annexes A.

---

```
#####
##### Regression par les MCO (France) #####
#####

#####
##### Regression avec toutes les variables de tous les individus #####
#####

#####
##### France #####
## Regression avec toutes les variables
model = lm(taux_horaires~age+situation+ secteur_trav
           +ancienneté+satisfaction+nivx_etude+
           profession_rec+heures_mois+Taille_entreprise
           , data=ewcs_2015_fra)
anova(model)
summary(model)

### Extraction des individu en temps plein
data_temps_plein = filter(ewcs_2015_fra, type_contrat == "Temps_plein")

#####
##### France #####
## Regression avec toutes les variables des individu en temps plein
model = lm(taux_horaires~age+situation+secteur_trav+Taille_entreprise
           +ancienneté+satisfaction+nivx_etude+profession_rec+heures_mois
           , data=data_temps_plein)
anova(model)
summary(model)

##### modèle constraint pour chow avec toute les variables #####
model_c = lm(salaire~age+situation+secteur_trav
              +Taille_entreprise+ancienneté+satisfaction+nivx_etude+
              profession_rec+heures_mois+Taille_entreprise, data=ewcs_2015_fra)
anova(model_c)
summary(model_c)
SCRC = sum(model_c$residuals^2)

#####
##### Objectif faire le test de CHOW
##### modèle non constraint avec toutes les variables #####
##### Homme #####
#####

model_nc1 = lm(salaire~age+situation+ secteur_trav
                +ancienneté+satisfaction+nivx_etude+
                profession+heures_mois+Taille_entreprise
                , data=ewcs_2015_fra, subset=(genre == "Homme"))
```

## Annexes A.

---

```
anova(model_nc1)
summary(model_nc1)

SCR1 = sum(model_nc1$residuals^2)

##### Femme #####
model_nc2 = lm(salaire~age+type_contrat+situation+ secteur_trav
               +ancienneté+satisfaction+nivx_etude+
               profession+heures_mois+Taille_entreprise
               , data=ewcs_2015_fra,subset=(genre == "Femme"))
anova(model_nc2)
summary(model_nc2)

SCR2 = sum(model_nc2$residuals^2)

SCRnc = SCR1 + SCR2

#### calcul de la statistique du test de Chow
dlnc <- (nrow(ewcs_2015_fra)-(2*length(coefficients(model_c))))
nb_contr <- length(coefficients(model_c))

Ftest <- ((SCRC-(SCR1+SCR2))/nb_contr)/((SCR1+SCR2)/dlnc)
pval <- 1-pf(Ftest,nb_contr,dlnc)
Chow_test <- c(Ftest,pval)
names(Chow_test) <- c("Stat F du test de Chow", "p-value")
round(Chow_test,3)

#####Tests de normalité (test de jacque bera)
jb.norm.test(model_c$residuals)
hist(resid(model_c),prob=TRUE,col = coul,main = "Fonction de densité estimée",xlab = "Résidu")
lines(density(resid(model_c) , na.rm = "TRUE"),lwd=3,col="cyan")
## Le 2
plot(model_c,which=1:6, col = coul)

##### Test d'hétéroscédasticité Femme #####
white <- lm(resid(model_c)^2~fitted(model_c)+I(fitted(model_c)^2), data=ewcs_2015_fra)
summary(white)

Ru2<- summary(white)$r.squared
LM <- nrow(ewcs_2015_fra)*Ru2
p.value <- 1-pchisq(LM, 2)
Test.het.White<-c(LM,p.value)
names(Test.het.White)<-c("Stat. LM", "p-value")
Test.het.White

summary(model_c)
library(xtable)
```

## Annexes A.

---

```
names(model_c)
a = resid(model_c)
residuals(model_c)
plot(x=a, main = "résidus du modèle", pch = 5, lwd = 3, col="red")

#####
##### Regression par les MCO avec toutes variables #####
##### des individus en temps plein #####
#####

### Extraction des individu en temps plein
data_temps_plein = subset(ewcs_2015_fra, type_contrat == "Temps_plein")

##### France #####
## Regression avec toutes les variables des individu en temps plein
model = lm(salaire~genre+age+situation+secteur_trav+Taille_entreprise
           +ancienneté+satisfaction+nivx_etude+profession_rec+heures_mois
           , data=data_temps_plein)
anova(model)
summary(model)

##### modèle constraint avec toute les variables #####
model_c = lm(salaire~genre+age+situation+ secteur_trav
              +ancienneté+satisfaction+nivx_etude+
              profession_rec+heures_mois+Taille_entreprise
              , data=data_temps_plein)
anova(model_c)
summary(model_c)

##### modèle non constraint avec toutes les variables #####
##### Homme #####
model_nc1 = lm(salaire~genre+age+type_contrat+situation+ secteur_trav
                +ancienneté+satisfaction+nivx_etude+
                profession+heures_mois+taux_horaires+Taille_entreprise
                , data=data_temps_plein, subset=(genre == "Homme"))
anova(model_nc1)
summary(model_nc1)

SCR1 = sum(model_nc1$residuals^2)

##### Femme #####
model_nc2 = lm(salaire~genre+age+type_contrat+situation+ secteur_trav
                +ancienneté+satisfaction+nivx_etude+profession+
                heures_mois+taux_horaires+Taille_entreprise
                , data=data_temps_plein,subset=(genre == "Femme"))
anova(model_nc2)
```

## Annexes A.

---

```
summary(model_nc2)

SCR2 = sum(model_nc2$residuals^2)

SCRnc = SCR1 + SCR2

##### calcul de la statistique du test de Chow
dlnc <- (nrow(data_temps_plein)-(2*length(coefficients(model_c))))
nb_contr <- length(coefficients(model_c))

Ftest <- ((SCRc-(SCR1+SCR2))/nb_contr)/((SCR1+SCR2)/dlnc)
pval <- 1-pf(Ftest,nb_contr,dlnc)
Chow_test <- c(Ftest,pval)
names(Chow_test) <- c("Stat F du test de Chow", "p-value")
Chow_test

##### Tests de normalité (test de jacque bera)
jb.norm.test(model_c$residuals)
shapiro.test(resid(model_c))

plot(model_c,which=1:6, col = coul)

##### Test d'hétéroscédasticité Femme #####
white <- lm(resid(model_c)^2~fitted(model_c)+I(fitted(model_c)^2), data=ewcs_2015_fra)
summary(white)

summary(model_c)
library(xtable)
names(model_c)
a = resid(model_c)
residuals(model_c)
plot(x=a, main = "résidus du modèle", pch = 5, lwd = 3, col="red")

hist(resid(model_c),prob=TRUE,col = "cornflowerblue",xlim =
      c(-19,19),main = "Fonction de densité estimée",xlab = "Résiduestimé",ylab = "Densité"

#####
#####
#####
#####
##### Regression avec l'équation de mincer pour la MCO #####
#####
#####
#####

### Tous les contrat en utilisant le taux horaires
model = lm(log(taux_horaires)~nivx_etude+ancienneté+I(ancienneté^2), data=ewcs_2015_fra)
anova(model)
```

## Annexes A.

---

```
summary(model)

### Extraction des individu en temps plein
data_temps_plein = filter(ewcs_2015_fra, type_contrat == "Temps_plein")

##### Temps plein #####
model_t_plein = lm(log(salaire)~nivx_etude+ancienneté+I(ancienneté^2), data=data_temps_plein)
anova(model_t_plein)
summary(model_t_plein)

##### modèle constraint #####
modelc_t_plein = lm(log(salaire)~nivx_etude+ancienneté+I(ancienneté^2), data=data_temps_plein)
anova(modelc_t_plein)
summary(modelc_t_plein)
SCRc = sum(modelc_t_plein$residuals^2)

##### modèle non constraint #####
##### Homme #####
modelnc_t_plein_1 = lm(log(salaire)~nivx_etude+ancienneté+I(ancienneté^2),
                       data=data_temps_plein,subset=(genre == "Homme"))
SCR_t_plein_1 = sum(modelnc_t_plein_1$residuals^2)

##### Femme #####
modelnc_t_plein_2 = lm(log(salaire)~nivx_etude+ancienneté+I(ancienneté^2),
                       data=data_temps_plein,subset=(genre == "Femme"))
SCR_t_plein_2 = sum(modelnc_t_plein_2$residuals^2)
SCRnc = SCR_t_plein_1 + SCR_t_plein_2

##### calcul de la statistique du test de Chow
dlnc <- (nrow(data_temps_plein)-(2*length(coefficients(modelc_t_plein))))
nb_contr <- length(coefficients(modelc_t_plein))

Ftest <- ((SCRc-(SCR_t_plein_1+SCR_t_plein_2))/nb_contr)/((SCR_t_plein_1+SCR_t_plein_2)/dlnc)
pval <- 1-pf(Ftest,nb_contr,dlnc)
Chow_test <- c(Ftest,pval)
names(Chow_test) <- c("Stat F du test de Chow", "p-value")
round(Chow_test, 3)

#####Tests de normalité (test de jacque bera)
jb.norm.test(modelc_t_plein$residuals)
shapiro.test(resid(modelc_t_plein))
plot(modelc_t_plein,which=1:6, col = coul)

##### Test d'hétéroscédasticité Femme #####
white <- lm(resid(modelc_t_plein)^2~fitted(modelc_t_plein)+I(fitted(modelc_t_plein)^2), data=data_temps_plein)
summary(white)
```

## Annexes A.

---

```
summary(modelc_t_ plein)
library(xtable)
names(modelc_t_ plein)
a = resid(modelc_t_ plein)
residuals(modelc_t_ plein)
plot(x=a, main = "résidus du modèle", pch = 5, lwd = 3, col="red")

hist(resid(modelc_t_ plein),prob=TRUE,col = "cornflowerblue",xlim =
      c(-19,19),main = "Fonction de densité estimée",xlab = "Résiduestimé",ylab = "Densité"

#####
##### Regression par les MCO (Allemagne) #####
#####

#####
##### Regression avec toutes les variables de tous les individus #####
#####

model = lm(salaire~age+situation+ secteur_trav
           +ancienneté+satisfaction+nivx_etude+profession_rec+
           heures_mois+Taille_entreprise, data=ewcs_2015_ger)

anova(model)
summary(model)

##### modèle constraint avec toute les variables #####
model_c = lm(salaire~age+situation+ secteur_trav
              +ancienneté+satisfaction+nivx_etude+
              profession+heures_mois+taux_horaires+Taille_entreprise
              , data=ewcs_2015_ger)
anova(model_c)
summary(model_c)
SRCc = sum(model_c$residuals^2)

#####Tests de normalité (test de jacque bera)
jb.norm.test(model_c$residuals)
hist(resid(model_c),prob=TRUE,col = coul,main = "Fonction de densité estimée",xlab = "Résidu
lines(density(resid(model_c), na.rm = "TRUE"),lwd=3,col="cyan")

#shapiro.test(resid(model_c))

plot(model_c,which=1:6, col = coul)

##### modèle non constraint avec toutes les variables #####
##### Homme #####
model_nc1 = lm(salaire~age+type_contrat+situation+secteur_trav+Taille_entreprise+
```

## Annexes A.

---

```
+ancienneté+satisfaction+nivx_etude+profession_rec+heures_mois  
,data=ewcs_2015_ger, subset=(genre == "Homme"))  
anova(model_nc1)  
summary(model_nc1)  
SCR1 = sum(model_nc1$residuals^2)  
  
##### Femme #####  
model_nc2 = lm(salaire~age+type_contrat+situation+ secteur_trav  
+ancienneté+satisfaction+nivx_etude+  
profession+heures_mois+Taille_entreprise  
, data=ewcs_2015_ger,subset=(genre == "Femme"))  
anova(model_nc2)  
summary(model_nc2)  
  
SCR2 = sum(model_nc2$residuals^2)  
SCRnc = SCR1 + SCR2  
  
#### calcul de la statistique du test de Chow  
dlnc <- (nrow(ewcs_2015_ger)-(2*length(coefficients(model_c))))  
nb_contr <- length(coefficients(model_c))  
  
Ftest <- ((SCRc-(SCR1+SCR2))/nb_contr)/((SCR1+SCR2)/dlnc)  
pval <- 1-pf(Ftest,nb_contr,dlnc)  
Chow_test <- c(Ftest,pval)  
names(Chow_test) <- c("Stat F du test de Chow", "p-value")  
round(Chow_test,3)  
  
#####Tests de normalité (test de jacque bera)  
jb.norm.test(model_c$residuals)  
shapiro.test(resid(model_c))  
hist(resid(model_c),prob=TRUE,col = coul,main = "Fonction de densité estimée",xlab = "Résidu")  
lines(density(resid(model_c)), na.rm = "TRUE"),lwd=3,col="cyan")  
  
plot(model_c,which=1:6, col = coul)  
  
##### Test d'hétéroscédasticité Femme #####  
white <- lm(resid(model_c)^2~fitted(model_c)+I(fitted(model_c)^2), data=ewcs_2015_fra)  
summary(white)  
  
summary(white)  
  
Ru2<- summary(white)$r.squared  
LM <- nrow(ewcs_2015_fra)*Ru2  
p.value <- 1-pchisq(LM, 2)  
Test.het.White<-c(LM,p.value)
```

## Annexes A.

---

```
names(Test.het.White)<-c("Stat. LM", "p-value")
round(Test.het.White,3)

summary(model_c)
library(xtable)
names(model_c)
a = resid(model_c)
residuals(model_c)
plot(x=a, main = "résidus du modèle", pch = 5, lwd = 3, col="red")

hist(resid(model_c),prob=TRUE,col = "cornflowerblue",xlim =
      c(-19,19),main = "Fonction de densité estimée",xlab = "Résiduestimé",ylab = "Densité"

#####
##### Regression par les MCQ avec toutes variables #####
##### des individus en temps plein #####
#####

### Extraction des individu en temps plein
data_temps_plein = filter(ewcs_2015_ger, type_contrat == "Temps_plein")

## Regression avec toutes les variables des individu en temps plein
model = lm(taux_horaires~genre+age+situation+ secteur_trav
           +ancienneté+satisfaction+nivx_etude+
           profession_rec+heures_mois
           , data=data_temps_plein)
anova(model)
summary(model)

##### modèle constraint avec toute les variables #####
model_c = lm(salaire~genre+age+type_contrat+situation+ secteur_trav
              +ancienneté+heure_par_semaine+satisfaction+nivx_etude+
              profession+heures_mois+taux_horaires
              , data=data_temps_plein)
anova(model_c)
summary(model_c)

##### modèle non constraint avec toutes les variables #####
##### Homme #####
model_nc1 = lm(salaire~age+type_contrat+situation+ secteur_trav
                +ancienneté+heure_par_semaine+satisfaction+nivx_etude+
                profession+heures_mois+taux_horaires
                , data=data_temps_plein, subset=(genre == "Homme"))
anova(model_nc1)
summary(model_nc1)
```

```

SCR1 = sum(model_nc1$residuals^2)

##### Femme #####
model_nc2 = lm(salaire~age+type_contrat+situation+ secteur_trav
               +ancienneté+Taille_entreprise+satisfaction+nivx_etude+
               profession+heures_mois+taux_horaires
               , data=data_temps_plein,subset=(genre == "Femme"))
anova(model_nc2)
summary(model_nc2)

SCR2 = sum(model_nc2$residuals^2)

SCRnc = SCR1 + SCR2

#### calcul de la statistique du test de Chow
dlnc <- (nrow(data_temps_plein)-(2*length(coefficients(model_c))))
nb_contr <- length(coefficients(model_c))

Ftest <- ((SCRc-(SCR1+SCR2))/nb_contr)/((SCR1+SCR2)/dlnc)
pval <- 1-pf(Ftest,nb_contr,dlnc)
Chow_test <- c(Ftest,pval)
names(Chow_test) <- c("Stat F du test de Chow", "p-value")
Chow_test

#####Tests de normalité (test de jacque bera)
jb.norm.test(model_c$residuals)
shapiro.test(resid(model_c))

plot(model_c,which=1:6, col = coul)

##### Test d'hétéroscéasticité Femme #####
white <- lm(resid(model_c)^2~fitted(model_c)+I(fitted(model_c)^2), data=ewcs_2015_fra)
summary(white)

summary(model_c)
library(xtable)
names(model_c)
a = resid(model_c)
residuals(model_c)
plot(x=a, main = "résidus du modèle", pch = 5, lwd = 3, col="red")

hist(resid(model_c),prob=TRUE,col = "cornflowerblue",xlim =
      c(-19,19),main = "Fonction de densité estimée",xlab = "Résiduestimé",ylab = "Densité"

#####
##### Regression avec l'équation de mincer pour la MCO #####

```

## Annexes A.

---

```
#####
##### Les individus avec tous les contrat #####
model = lm(log(taux_horaires)~nivx_etude+ancienneté+I(ancienneté^2), data=ewcs_2015_ger)
anova(model)
summary(model)

##### modèle constraint #####
modelc = lm(log(salaire)~nivx_etude+ancienneté+I(ancienneté^2), data=ewcs_2015_ger)
anova(modelc)
summary(modelc)
SCRc = sum(modelc$residuals^2)

##### modèle non constraint #####
##### Homme #####
modelnc_1 = lm(log(salaire)~nivx_etude+ancienneté+I(ancienneté^2),
               data=ewcs_2015_ger,subset=(genre == "Homme"))
SCR_1 = sum(modelnc_1$residuals^2)

##### Femme #####
modelnc_2 = lm(log(salaire)~nivx_etude+ancienneté+I(ancienneté^2),
               data=ewcs_2015_ger,subset=(genre == "Femme"))
SCR_2 = sum(modelnc_t_partiel_2$residuals^2)

SCRnc = SCR_1 + SCR_2

#### calcul de la statistique du test de Chow
dlnc <- (nrow(ewcs_2015_ger)-(2*length(coefficients(modelc))))
nb_contr <- length(coefficients(modelc))

Ftest <- ((SCRc-(SCR_1+SCR_2))/nb_contr)/((SCR_1+SCR_2)/dlnc)
pval <- 1-pf(Ftest,nb_contr,dlnc)
Chow_test <- c(Ftest,pval)
names(Chow_test) <- c("Stat F du test de Chow", "p-value")
Chow_test

#####Tests de normalité (test de jacque bera)
jb.norm.test(modelc$residuals)
shapiro.test(resid(modelc))

plot(modelc,which=1:6, col = coul)

##### Test d'hétéroscédasticité Femme #####
white <- lm(resid(modelc)^2~fitted(modelc)+
            I(fitted(modelc)^2), data=ewcs_2015_ger)
summary(white)
```

## Annexes A.

---

```
summary(modelc)
library(xtable)
names(modelc)
a = resid(modelc)
residuals(modelc)
plot(x=a, main = "résidus du modèle",
      pch = 5, lwd = 3, col="red")

hist(resid(modelc),prob=TRUE,
      col = "cornflowerblue",xlim =c(-19,19),
      main = "Fonction de densité estimée",
      xlab = "Résiduestimé",ylab = "Densité")

### Extraction des individu en temps plein
data_temps_plein = filter(ewcs_2015_ger, type_contrat == "Temps_plein")
model= lm(log(salaire)~nivx_etude+ancienneté+I(ancienneté^2), data=data_temps_plein)
anova(model)
summary(model)

##### Les individus en temps plein #####
#### modèle constraint #####
modelc_t_plein = lm(log(salaire)~nivx_etude+ancienneté+I(ancienneté^2), data=data_temps_plein)

anova(modelc_t_plein)
summary(modelc_t_plein)
SCRc = sum(modelc_t_plein$residuals^2)

#### modèle non constraint #####
##### Homme #####
modelnc_t_plein_1 = lm(log(taux_horaires)~nivx_etude+ancienneté+I(ancienneté^2),
                       data=data_temps_plein,subset=(genre == "Homme"))
SCR_t_plein_1 = sum(modelnc_t_plein_1$residuals^2)

##### Femme #####
modelnc_t_plein_2 = lm(log(salaire)~nivx_etude+ancienneté+I(ancienneté^2),
                       data=data_temps_plein,subset=(genre == "Femme"))

SCR_t_plein_2 = sum(modelnc_t_plein_2$residuals^2)
SCRnc = SCR_t_plein_1 + SCR_t_plein_2

#### calcul de la statistique du test de Chow
dlnc <- (nrow(data_temps_plein)-(2*length(coefficients(modelc_t_plein))))
nb_contr <- length(coefficients(modelc_t_plein))

Ftest <- ((SCRc-(SCR_t_plein_1+SCR_t_plein_2))/nb_contr)/((SCR_t_plein_1+SCR_t_plein_2)/dlnc)
pval <- 1-pf(Ftest,nb_contr,dlnc)
Chow_test <- c(Ftest,pval)
names(Chow_test) <- c("Stat F du test de Chow", "p-value")
```

## Annexes A.

---

Chow\_test

```
#####Tests de normalité (test de jacque bera)
jb.norm.test(modelc_t_ plein$residuals)
shapiro.test(resid(modelc_t_ plein))

plot(modelc_t_ plein,which=1:6, col = coul)

#####Test d'hétéroscédasticité Femme #####
white <- lm(resid(modelc_t_ plein)^2~fitted(modelc_t_ plein)+I(fitted(modelc_t_ plein)^2), data)
summary(white)

summary(modelc_t_ plein)
library(xtable)
names(modelc_t_ plein)
a = resid(modelc_t_ plein)
residuals(modelc_t_ plein)
plot(x=a, main = "résidus du modèle", pch = 5, lwd = 3, col="red")

hist(resid(modelc_t_ plein),prob=TRUE,col = "cornflowerblue",xlim =
      c(-19,19),main = "Fonction de densité estimée",xlab = "Résiduestimé",ylab = "Densité"

##### Fin du code #####
```

Le tableau ci dessous montre de la gauche vers la droite,les trois première composante, qualité de représentation et la contribution.

Tableau : Composante principale, qualité de représentation et contribution

---

## BIBLIOGRAPHIE

- [1] Meurs Dominique, Ponthieux Sophie. (2000), *Une mesure de la discrimination dans l'écart de salaire entre hommes et femmes* , in : Economie et statistique, n°337-338,Jeunes : l'âge des indépendances. pp. 135-158 ;
- [2] Meurs Dominique, Ponthieux Sophie. (2002), *L'écart des salaires entre les femmes et les hommes peut-il encore baisser ?* , in : ÉCONOMIE ET STATISTIQUE N° 398-399, 2006, Jeunes : l'âge des indépendances. pp. 135-158 ;
- [3] Ricco RAKOTOMALALA(2018), *Analyse Factorielle de Données Mixtes AFDM*,Principes et pratique de l'AFDM, Université Lumière Lyon 2.
- [4] J. PAGÈS, *Analyse factorielle de données mixtes*, Revue de statistique appliquée, tome 52, no 4 (2004), p. 93-111.
- [5] B. ESCOFIER(2018), *Traitemet simultané de variables qualitatives et quantitatives en analyse factorielle*, Les cahiers de l'analyse des données, tome 4, no 2 (1979),p. 137-146.
- [6] Jérôme Pagès(2018), *ANALYSE FACTORIELLE DE DONNÉES MIXTES : PRINCIPE ET EXEMPLE D'APPLICATION*, Laboratoire de mathématiques appliquées,Agrocampus, 35042 Rennes cedex.
- [7] <https://www.researchgate.net/publication/264038510>
- [8] [http://www.numdam.org/item/?id=CAD\\_1979\\_\\_4\\_2\\_137\\_0](http://www.numdam.org/item/?id=CAD_1979__4_2_137_0)
- [9] [http://www.numdam.org/item/?id=RSA\\_2004\\_\\_52\\_4\\_93\\_0](http://www.numdam.org/item/?id=RSA_2004__52_4_93_0)
- [10] [tutoriels-data-mining.blogspot.com](http://tutoriels-data-mining.blogspot.com)
- [11] <http://sthdna.com/english/articles/31-principal-component-methods-in-r-practical-guide/115-fAMD-factor-analysis-of-mixed-data-in-r-essentials/>
- [12] <http://rdrr.io/github/husson/FactoMineR/src/R/FAMD.R>
- [13] <https://husson.github.io/data.html>