

UNIVERSITÉ EVRY

FACULTÉ DE DROIT, ECONOMIE ET GESTION

TITRE

**L'IMPACT DE L'EDUCATION SUR LE MARCHE
DU TRAVAIL FRANÇAIS**

**UNE ANALYSE ECONOMETRIQUE DU LIEN ENTRE NIVEAU D'EDUCATION,
TAUX DE CHOMAGE ET REVENU**

NOM, Prénom

NOM, Prénom

NOM, Prénom

Groupe de TD No

Chargé de TD M. ... ou Mme

LICENCE 3ÈME ANNÉE – Sciences Économiques et de Gestion

Semestre 2

Introduction à l'économétrie

Année universitaire 2023-2024

Table des matières

1.	Introduction.....	1
2.	Présentation des données.....	1
3.	Analyse exploratoire des données	2
3.1.	Importation de jeu de données :	2
3.2.	Statistique descriptive des données :	3
4.	Visualisation des données :.....	4
4.2.	Visualisation des données catégorielles :.....	4
4.3.	Visualisation des variables numériques :	8
4.3.	Visualisation des analyses croisées :.....	11
5.	Construction des modèles :.....	15
5.1.	Création du modèle de régression linéaire simple :.....	15
5.1.1.	Etude de relation entre le taux d'insertion professionnel et poids de spécialité :15	15
5.1..	Etude de relation entre le salaire brute et le taux de chômage :	18
5.2.	Test d'homogénéité des modèles :.....	19
5.3.	Création du modèle de régression linéaire multiple :.....	21
5.3.1.	Salaire brute en fnct du taux de chômage, taux d'insertion, poids de discipline 21	21
5.3.2.	Etude de salaire brut en fonction de la spécialité :.....	22
5.3.3.	Création de modèle multiple globale :.....	24
6.	Performance du modèle :.....	25
7.	Vérification des hypothèses :.....	28
8.	Recommandations :.....	29
9.	Conclusion.....	30
10.	Bibliographie.....	33
11.	Source.....	34

1. Introduction

Lorsque nous examinons les rouages de la société française, certains éléments cruciaux se dégagent : l'éducation, le taux de chômage et le revenu. Ces trois piliers constituent des facettes essentielles de la vie des citoyens, influençant leurs trajectoires personnelles, leurs aspirations et leur bien-être économique. En France, l'Institut national de la statistique et des études économiques (INSEE) joue un rôle clé en mettant à notre disposition une vaste quantité de données qui nous permettent d'explorer en profondeur ces domaines.

Partant sur ces informations, et au cœur de notre démarche se trouve une question fondamentale

Quels sont les liens qui unissent l'éducation, le taux de chômage et le revenu en France ?

Cette problématique centrale guidera notre analyse économétrique, cherchant à décrypter les interactions complexes entre ces aspects cruciaux de la vie des individus, en utilisant des outils statistiques de pointe.

A fin de répondre à cette question, nous formulons plusieurs hypothèses de travail :

- **Hypothèse 1 :** Nous presupposons que le niveau de diplôme est étroitement lié au salaire brut annuel estimé, avec des diplômés de niveaux supérieurs gagnant en moyenne des salaires plus élevés.
- **Hypothèse 2 :** Nous considérons que le taux de chômage national peut influencer les opportunités d'emploi, avec un taux de chômage plus élevé réduisant les perspectives d'emploi pour les individus.
- **Hypothèse 3 :** Nous envisageons que le choix du domaine d'études puisse avoir un impact sur les revenus, certains domaines étant potentiellement plus lucratifs que d'autres.

Cette étude s'attache à simplifier les relations complexes entre l'éducation, le taux de chômage et le revenu en France. En utilisant des méthodes économétriques, nous chercherons à quantifier ces relations et à identifier leurs implications importantes pour les politiques éducatives et de l'emploi.

2. Présentation des données

Dans le cadre de notre étude sur les relations entre l'éducation, le taux de chômage et le revenu en France, nous disposons d'une base de données riche et diversifiée provenant de l'Institut national de la statistique et des études économiques (INSEE). Ces données couvrent plusieurs années et sont organisées en plusieurs variables, chacune apportant une perspective unique sur

la situation socio-économique en France. Voici une brève présentation des principales variables disponibles dans notre jeu de données :

- **Année** : Cette variable nous indique l'année à laquelle les données se rapportent. Elle nous permet de suivre les évolutions au fil du temps.
- **Diplôme** : Cette variable identifie le niveau de diplôme ou d'éducation atteint par les individus, allant du niveau le plus bas au plus élevé.
- **Situation** : Il s'agit du statut professionnel des individus, indiquant s'ils sont employés, chômeurs, étudiants, retraités, etc.
- **Code du Domaine** : Un code numérique désignant un domaine d'études ou de travail spécifique.
- **Domaine** : Le nom du domaine d'études ou de travail correspondant au code du domaine.
- **Code de la Discipline** : Un code numérique qui spécifie une discipline particulière au sein du domaine.
- **Salaire Brut Annuel Estimé** : Elle nous fournit des informations sur le salaire brut annuel moyen estimé pour les individus.
- **Taux de Chômage National** : C'est le taux de chômage au niveau national, qui mesure la proportion de la population active sans emploi.
- **Taux d'Insertion** : Le taux d'insertion mesure la proportion d'individus récemment diplômés qui ont trouvé un emploi dans leur domaine d'études ou de formation.

Ces données riches et variées nous offrent la possibilité d'explorer des questions essentielles liées à l'éducation, à l'emploi et aux revenus en France, en utilisant des méthodes d'analyse statistique avancées. Nous allons maintenant plonger plus en profondeur pour analyser et interpréter ces données dans le cadre de notre recherche.

3. Analyse exploratoire des données

3.1. Importation de jeu de données :

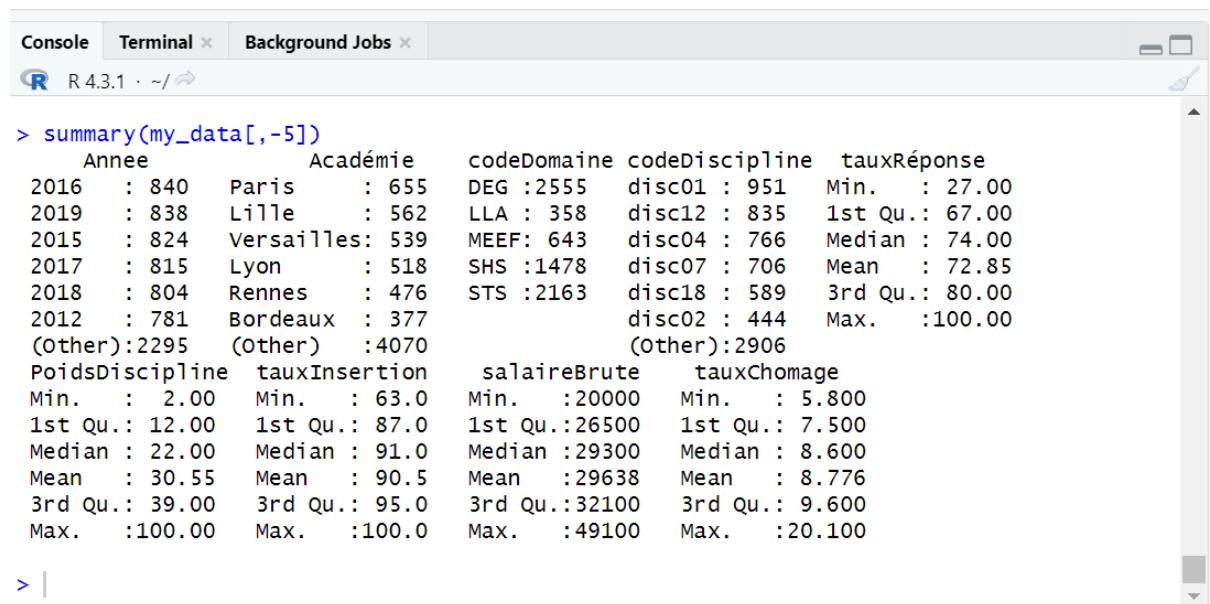
Nous avons utilisé pendant la réalisation de cette étude le logiciel « R-Studio », qui est un logiciel très puissant dans l'analyse statistique, en appuyant sur plusieurs Library et fonctions montrées dans **l'annexe 01**.

Nous avons nommé la base de données sous le nom « Master » lors de l'importation de cette dernière qui était sous forme d'un document CSV.

3.2. Statistique descriptive des données :

Après l'importation de notre jeu de donnée, nous avons utilisé la fonction summary (), pour montrer les différents résultats statistiques de la base que nous dispose, les résultats ont été comme suit :

Figure 01 : Résumé statistique de jeu de données.



```
Console Terminal × Background Jobs ×
R 4.3.1 · ~/ ↗

> summary(my_data[, -5])
   Année      Académie    codeDomaine codeDiscipline   tauxRéponse
2016   : 840    Paris     : 655      DEG :2555    disc01 : 951    Min.   : 27.00
2019   : 838    Lille     : 562      LLA : 358    disc12 : 835    1st Qu.: 67.00
2015   : 824    Versailles: 539      MEEF: 643    disc04 : 766    Median : 74.00
2017   : 815    Lyon      : 518      SHS :1478    disc07 : 706    Mean   : 72.85
2018   : 804    Rennes    : 476      STS :2163    disc18 : 589    3rd Qu.: 80.00
2012   : 781    Bordeaux  : 377          disc02 : 444    Max.   :100.00
(Other):2295 (Other)   :4070          (Other):2906
PoidsDiscipline   tauxInsertion    salaireBrute   tauxChomage
Min.   : 2.00    Min.   : 63.0    Min.   :20000    Min.   : 5.800
1st Qu.: 12.00   1st Qu.: 87.0    1st Qu.:26500   1st Qu.: 7.500
Median : 22.00   Median : 91.0    Median :29300    Median : 8.600
Mean   : 30.55   Mean   : 90.5    Mean   :29638    Mean   : 8.776
3rd Qu.: 39.00   3rd Qu.: 95.0    3rd Qu.:32100   3rd Qu.: 9.600
Max.   :100.00   Max.   :100.0    Max.   :49100    Max.   :20.100
> |
```

Nous constatons que (On prend que les variables d'intérêts) :

- **Année** : La répartition inégale des données par année suggère que certaines années ont été davantage étudiées que d'autres. Cela peut être lié à des facteurs économiques, politiques ou de recherche, nos données réparties sur plusieurs années, allant de 2011 à 2015.
- **SalaireBrut** : Les statistiques indiquent que le salaire brut annuel estimé varie considérablement, de 21 400 € à 43 900 €. Cela souligne les disparités de revenus économiques dans l'échantillon.
- **TauxChomage** : La variation du taux de chômage, de 8,6 % à 10,1 %, peut avoir des implications sur le marché du travail et l'économie en général. Une analyse plus poussée pourrait explorer les causes de ces variations.
- **TauxInsertion** : Le taux d'insertion, qui indique la réussite des diplômés à trouver un emploi dans leur domaine d'études, montre des variations significatives. Une exploration approfondie pourrait révéler des tendances liées à l'éducation et à l'emploi.

- **codeDom** : Les différentes catégories de "codeDom" représentent probablement des domaines d'études ou de travail. Les codes "STS" et "DEG" sont les plus fréquents, ce qui pourrait refléter des domaines d'études dominants ou des secteurs d'emploi prédominants dans l'échantillon.

4. Visualisation des données :

Après la statistique descriptive de données, nous allons visualiser les profits qui est la variable d'intérêt de notre étude, on va la visualiser avec la variable state pour voir quelle région qui possède le plus de gain.

4.2. Visualisation des données catégorielles :

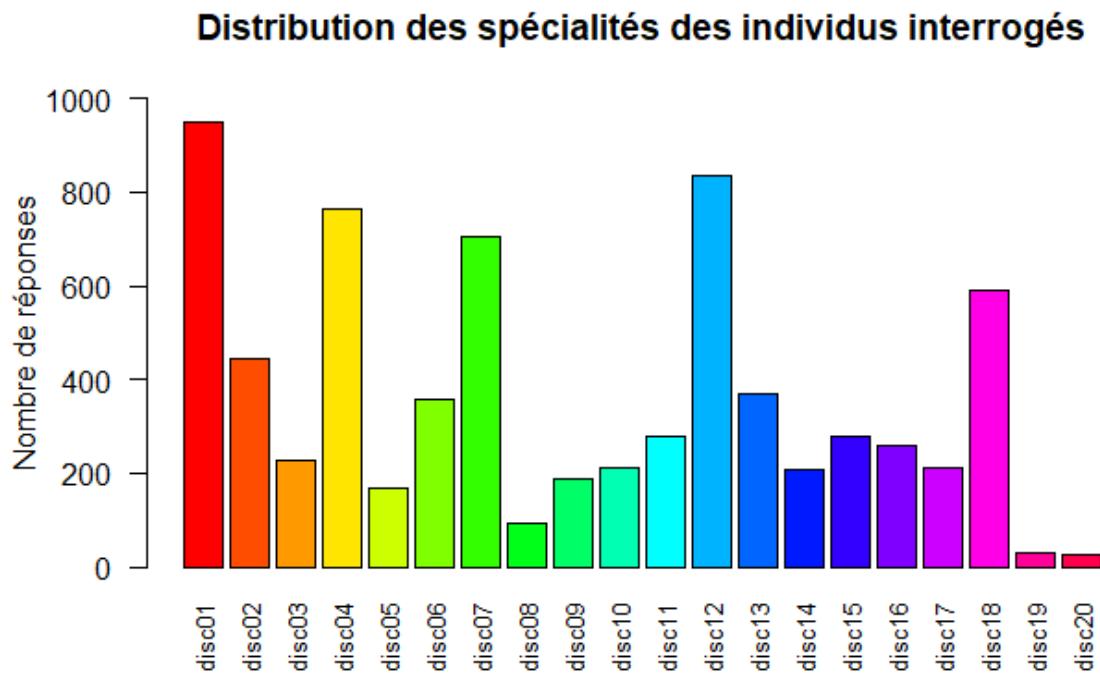
- **Code de discipline :**

La variable code de discipline montre les 20 spécialités étudiées au sein des différentes universités en France, nous différencions 20 disciplines différentes comme montre le tableau ci-dessus :

Table 1 : libellé des spécialités selon le code de domaine.

Code de la discipline	Libellé Discipline
disc01	Ensemble Formations Juridiques, Economique et De Gestion
disc02	Droit
disc03	Economie
disc04	Gestion
disc05	Autres Formations Juridiques, Economique et De Gestion
disc06	Lettres, Langues, Arts
disc07	Ensemble Sciences Humaines Et Sociales
disc08	Histoire-Geographique
disc09	Psychologie
disc10	Information Communication
disc11	Autres Sciences Humaines Et Sociales
disc12	Ensemble Sciences, Technologies Et Sante
disc13	Sciences De La Vie Et De La Terre
disc14	Sciences Fondamentales
disc15	Sciences De L'ingenieur
disc16	Informatique
disc17	Autres Sciences, Technologies Et Sante
disc18	Masters Enseignement
disc19	Masters Enseignement : Premier Degre
disc20	Masters Enseignement : Second Degre, CPE...

Figure 1 : Distribution des spécialités des individus.



Les données montrent une répartition diversifiée des spécialités d'étude parmi les individus de notre échantillon. L'analyse révèle que la spécialité "Droit" compte le plus grand nombre d'individus, avec un effectif de 444. Elle est suivie de près par la spécialité "Gestion" avec 766 individus, et "Masters enseignement" avec 589 individus. Les sciences sont également bien représentées, avec 368 individus dans "Sciences de la vie et de la terre" et 280 individus dans "Sciences de l'ingénieur".

Il est intéressant de noter que certaines catégories, comme "formations juridiques, économiques et de gestion" avec 168 individus, "Autres sciences humaines et sociales" avec 279 individus, et "Autres sciences, technologies et santé" avec 213 individus, représentent également des parts importantes de l'échantillon.

Ces résultats nous donnent un aperçu des préférences de spécialité d'étude au sein de notre échantillon, ce qui peut être précieux pour orienter nos analyses et nos décisions futures.

- **Code de domaine :**

La variable code de domaine montre les 05 domaines essentiels des individus interrogés dans cette étude, le tableau ci-dessus montre les libellés des codes domaines :

Table 2 : Libellé des codes du domaine.

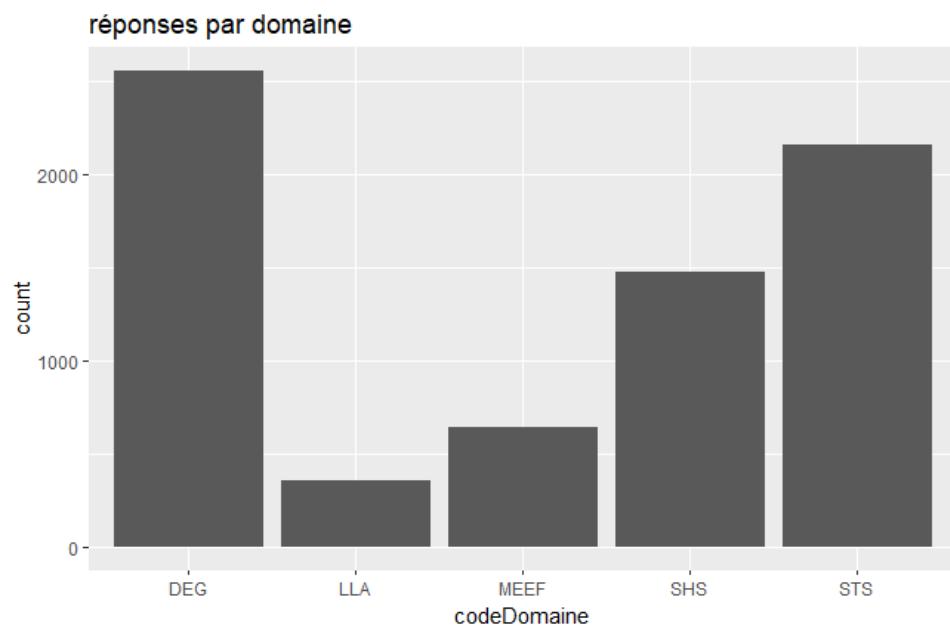
Code du domaine	Libellé du domaine	Nombre de réponses
DEG	Droit et gestion	2555
SHS	Sciences humaine et sociales	2163
STS	Sciences technologiques et santé	1478
LLA	Lettres, langues et arts	358
MEEF	Master enseignement	643

Nos données révèlent une répartition significative des individus parmi les différents domaines de spécialité. Le domaine Droit et gestion compte le plus grand nombre d'individus, avec un effectif de 2 555. Il est suivi par le domaine Sciences technologiques et santé avec 2 163 individus, et Sciences humaines et sociales avec 1 478 individus.

D'autre part, le domaine Lettres, langues et arts représente également une part substantielle de l'échantillon, avec 358 individus, suivi par le domaine Master enseignement avec 643 individus.

Ces résultats mettent en évidence la diversité des domaines de spécialité parmi les individus de notre échantillon. Ils suggèrent également que les domaines liés aux sciences et au droit et à la gestion sont particulièrement populaires parmi nos répondants. Ces informations peuvent être précieuses pour comprendre les tendances en matière d'éducation et d'orientation professionnelle au sein de notre population d'étude.

Figure 2 : Distribution des domaines d'études des individus.



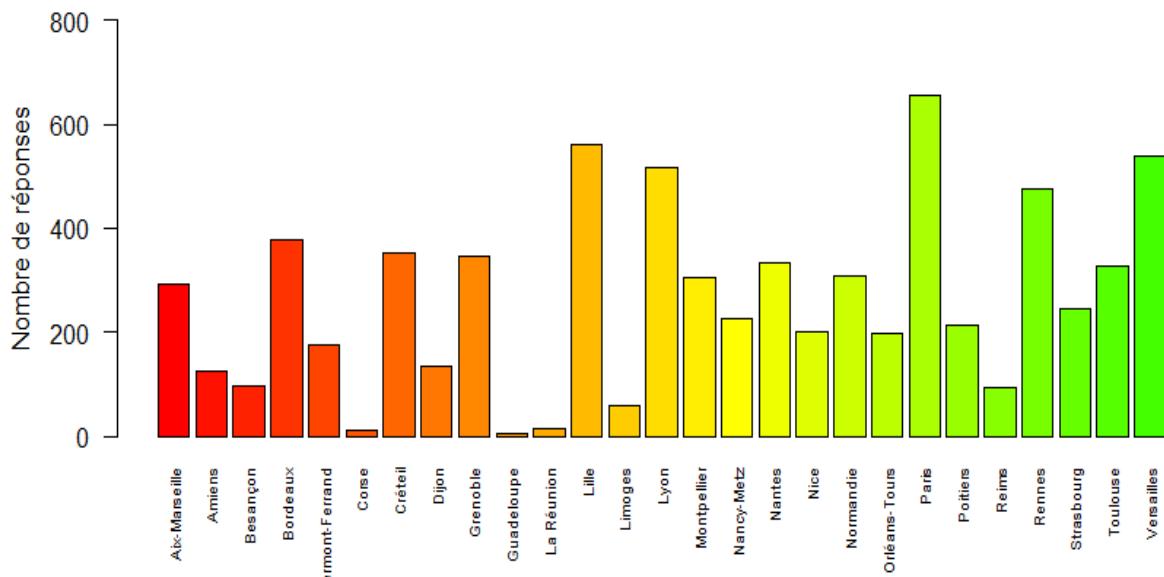
- Académie :

La variable académie représente les différentes régions de France et le nombre des réponses des individus sur la base des académies universitaires sur ces régions,

Régions	Nombre de réponses	Régions	Nombre de réponses
Aix-Marseille	291	Limoges	58
Amiens	126	Lyon	518
Besançon	98	Montpellier	306
Bordeaux	377	Nancy-Metz	227
Clermont-Ferrand	176	Nantes	333
Corse	11	Nice	202
Créteil	353	Normandie	309
Dijon	134	Orléans-Tours	198
Grenoble	347	Paris	655
Guadeloupe	6	Poitiers	214
La Réunion	16	Reims	93
Lille	562	Rennes	476
Strasbourg	245	Versailles	539
Toulouse	327		

Figure 3 : Distribution des individus selon les régions.

Distribution de nombre de réponses des individus selon les régions de france



Les résultats de notre étude révèlent une distribution variée du nombre de réponses collectées dans différentes régions académiques. Parmi les régions académiques, Paris se distingue clairement avec le plus grand nombre de réponses, totalisant 655 répondants. Cette forte participation suggère un niveau d'engagement élevé dans notre étude.

Les régions académiques de Lille et Versailles suivent de près, avec respectivement 562 et 539 réponses, ce qui indique également un taux de réponse significatif.

D'autres régions, telles que Rennes avec 476 réponses et Grenoble avec 347 réponses, montrent également une participation importante, ce qui témoigne de l'intérêt de ces régions pour notre étude.

En revanche, certaines régions académiques, comme Guadeloupe avec seulement 6 réponses et Corse avec 11 réponses, ont enregistré un nombre de réponses notablement plus faible. Cela peut être dû à divers facteurs, notamment la taille de la population locale ou la disponibilité des répondants.

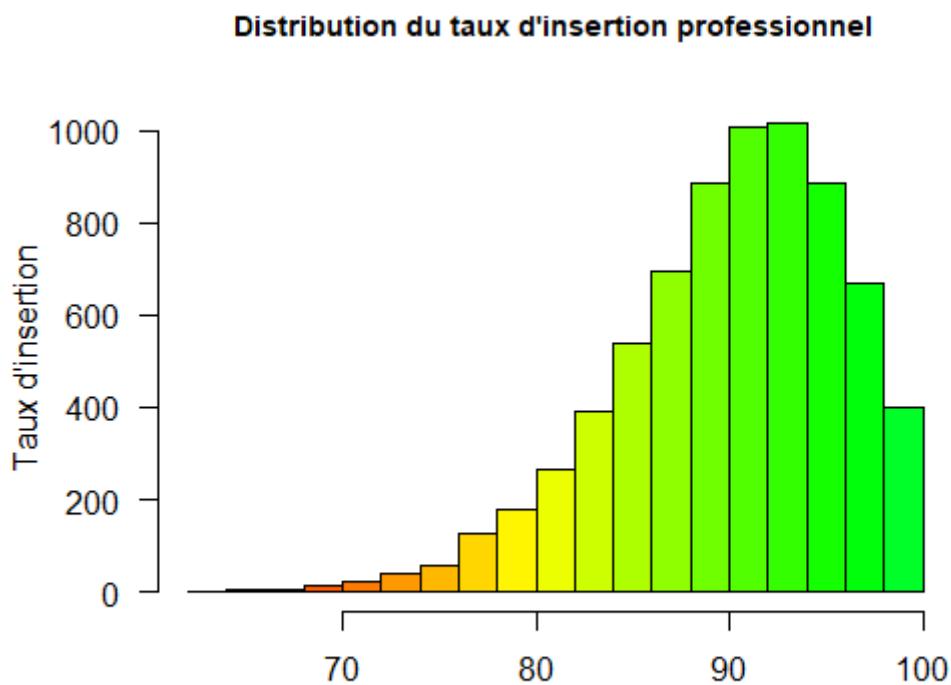
4.3. Visualisation des variables numériques :

- **Taux d'insertion :**

En examinant les statistiques descriptives de la variable de taux d'insertion professionnelle pour les universités de différentes régions françaises sur la période de 2010 à 2019, nous pouvons dégager certaines tendances. La gamme de taux d'insertion est variée, avec un taux minimum de 63 %, indiquant que certaines universités ont rencontré des défis en matière d'employabilité. Cependant, la majorité des universités affichent des taux d'insertion professionnelle supérieurs à 87 %, soulignant une employabilité généralement solide pour la plupart des diplômés.

La médiane de 91 % révèle que la moitié des universités ont atteint un taux d'insertion professionnel supérieur à ce seuil. Néanmoins, la moyenne légèrement inférieure à la médiane suggère que quelques valeurs plus basses ont tiré la moyenne vers le bas. Un quart des universités ont réussi à atteindre un taux d'insertion professionnel de 95 % ou plus, démontrant un excellent résultat en termes d'employabilité. Enfin, le taux maximum de 100 % met en évidence le potentiel d'obtenir une employabilité parfaite. Ces résultats nous incitent à poursuivre nos efforts pour améliorer l'employabilité des diplômés, en particulier dans les universités qui rencontrent des défis.

Figure 4 : Distribution du taux d'insertion des universités dans toutes les régions de France.



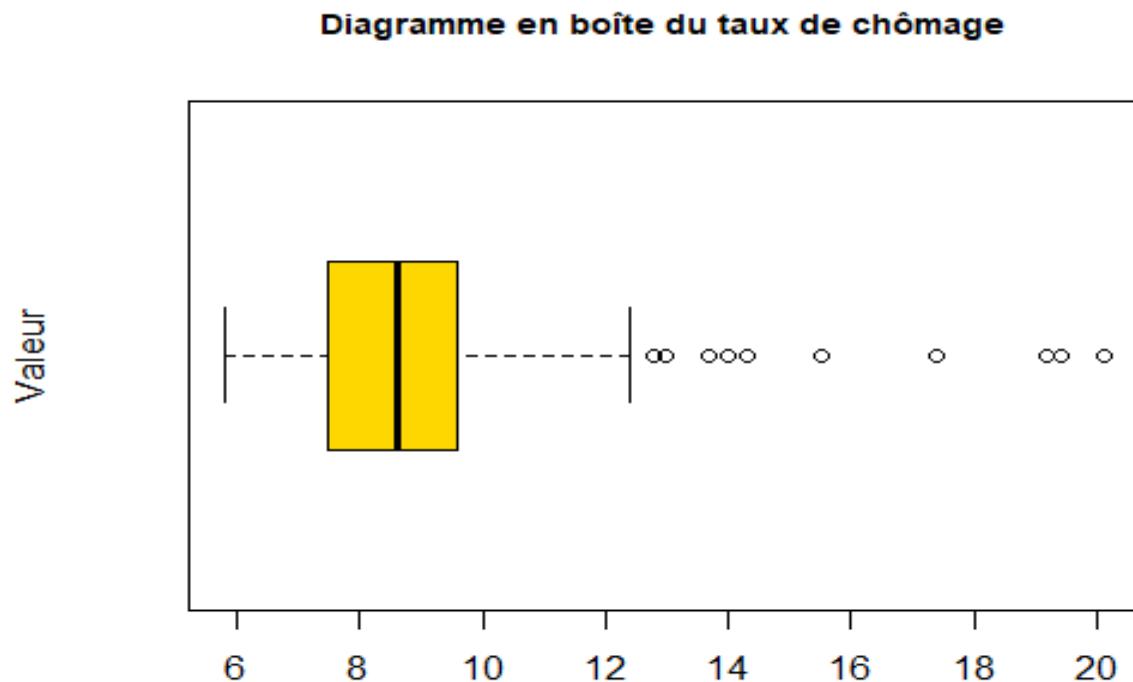
- **Taux de chômage :**

En analysant les statistiques descriptives du taux de chômage dans différentes régions françaises pour la période de 2010 à 2019, nous pouvons identifier certaines tendances significatives. Le taux de chômage le plus bas enregistré au cours de cette période est de 5,8 %, tandis que le premier quartile se situe à 7,5 %. Cela suggère que, pour 25 % des régions, le taux de chômage était inférieur à 7,5 %, indiquant une relative stabilité sur le marché de l'emploi dans ces régions.

La médiane du taux de chômage est de 8,6 %, ce qui signifie que la moitié des régions avaient un taux de chômage inférieur à ce seuil, tandis que l'autre moitié avait des taux supérieurs. La moyenne, légèrement plus élevée à 8,776 %, est influencée par quelques valeurs plus élevées, indiquant une certaine variation dans le taux de chômage à travers les régions.

Le troisième quartile, à 9,6 %, révèle que pour 25 % des régions, le taux de chômage était supérieur à ce seuil. Enfin, le taux de chômage le plus élevé enregistré au cours de cette période est de 20,1 %. Il est important de noter que des taux de chômage élevés peuvent indiquer des difficultés sur le marché de l'emploi, ce qui peut nécessiter des politiques et des mesures pour stimuler la création d'emplois et réduire le chômage.

Figure 5 : Distribution du taux de chômage régionales en France.



- **Salaire Brute :**

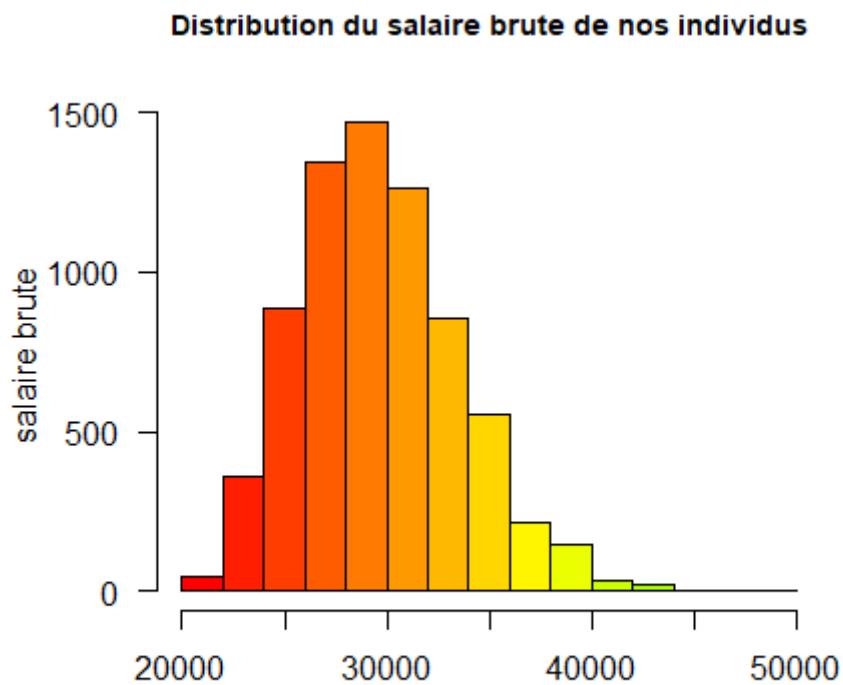
Les statistiques descriptives du salaire brut annuel estimé des individus révèlent des informations importantes sur la distribution des salaires au sein de notre échantillon. Le salaire brut annuel le plus bas enregistré est de 20 000 euros, ce qui représente le revenu minimum parmi les individus inclus dans notre étude.

Le premier quartile est de 26 500 euros, ce qui signifie que 25 % des individus gagnent un salaire inférieur à cette valeur. La médiane, quant à elle, est de 29 300 euros, indiquant que la moitié de notre échantillon gagne moins de ce montant, tandis que l'autre moitié gagne plus.

La moyenne du salaire brut annuel estimé est de 29 638 euros, ce qui donne une idée générale de la rémunération moyenne dans notre échantillon. Le troisième quartile est de 32 100 euros, signalant que 25 % des individus gagnent plus de cette somme.

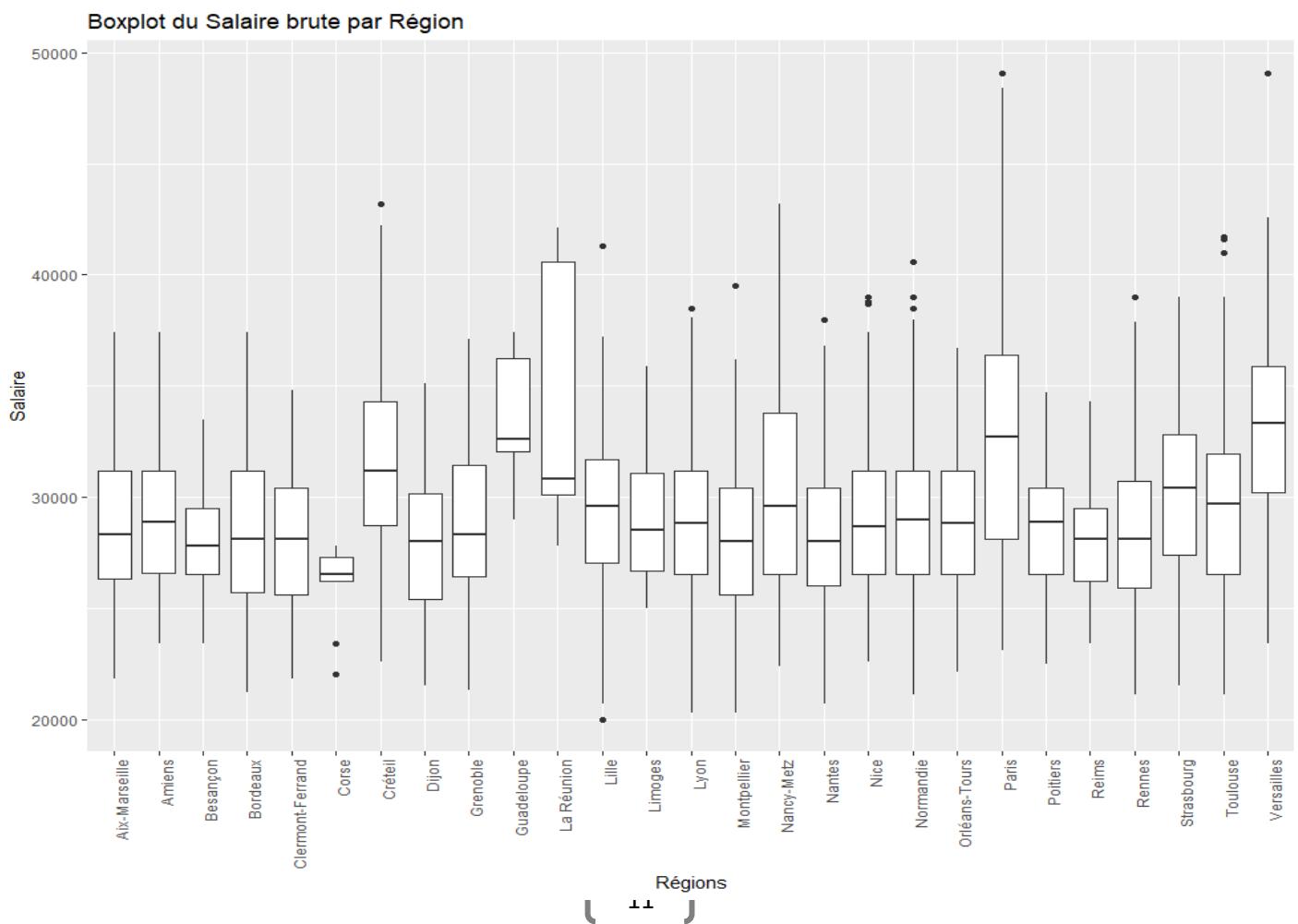
Le salaire brut annuel le plus élevé dans notre échantillon est de 49 100 euros, démontrant la variation des revenus parmi les individus de notre étude.

Figure 6 : Distribution des salaires brutes des individus.



4.3. Visualisation des analyses croisées :

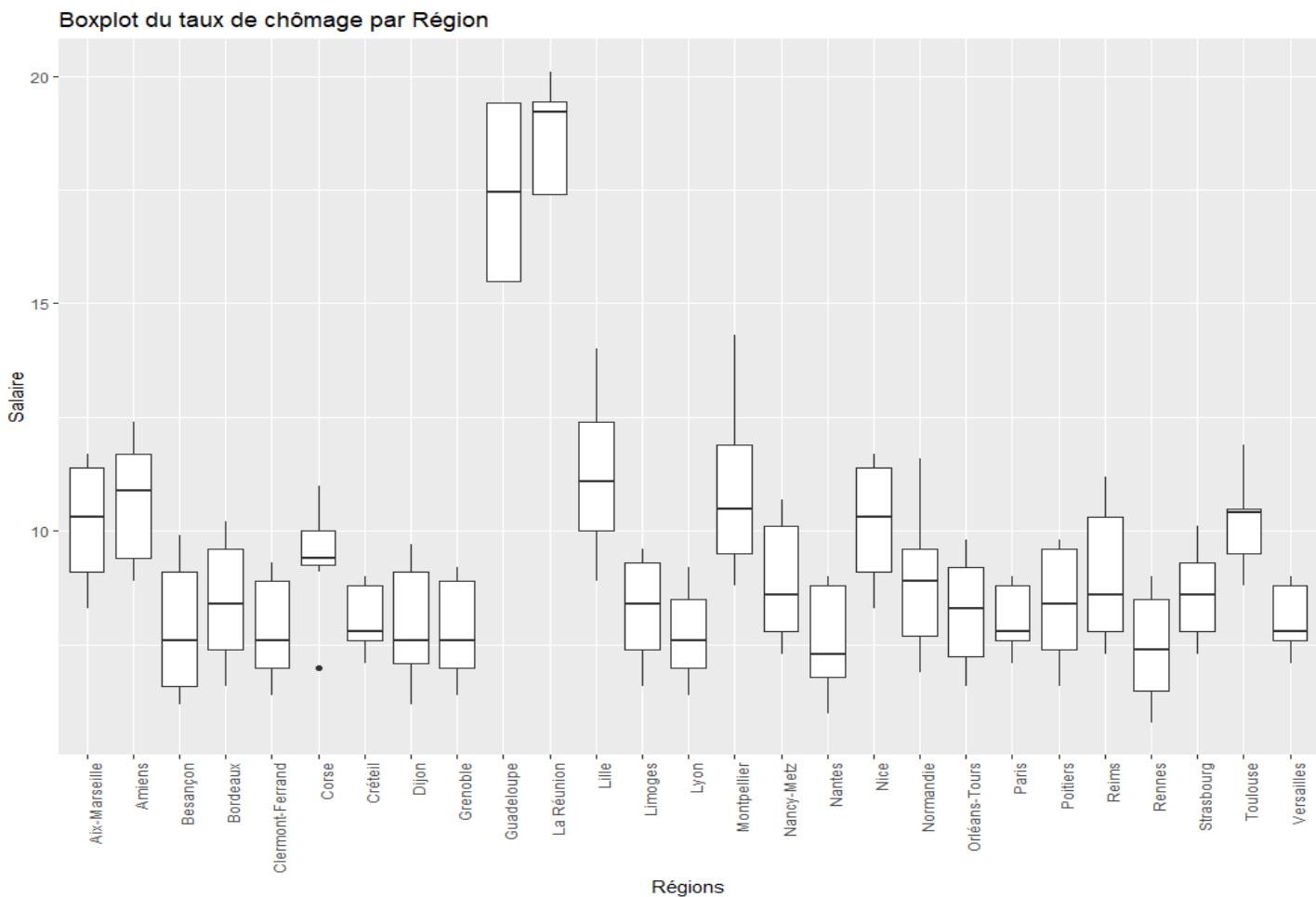
- Analyse des distributions du salaire brute avec les différentes régions françaises :



Nous constatons que la moyenne des salaires brutes en France diffèrent d'une région à une autre, les régions du sud semblent avoir un salaire brute équivalent qui proche du 26000 euro, et c'est pareil pour les régions de l'est et ouest comme Lyon, Normandie orléans.

La meilleure moyenne des salaires brutes à été marqué dans la région parisienne et ile de France avec un salaire brute moyen de 35000 euros, et c'est le cas pour la région du Guadeloupe et la réunion.

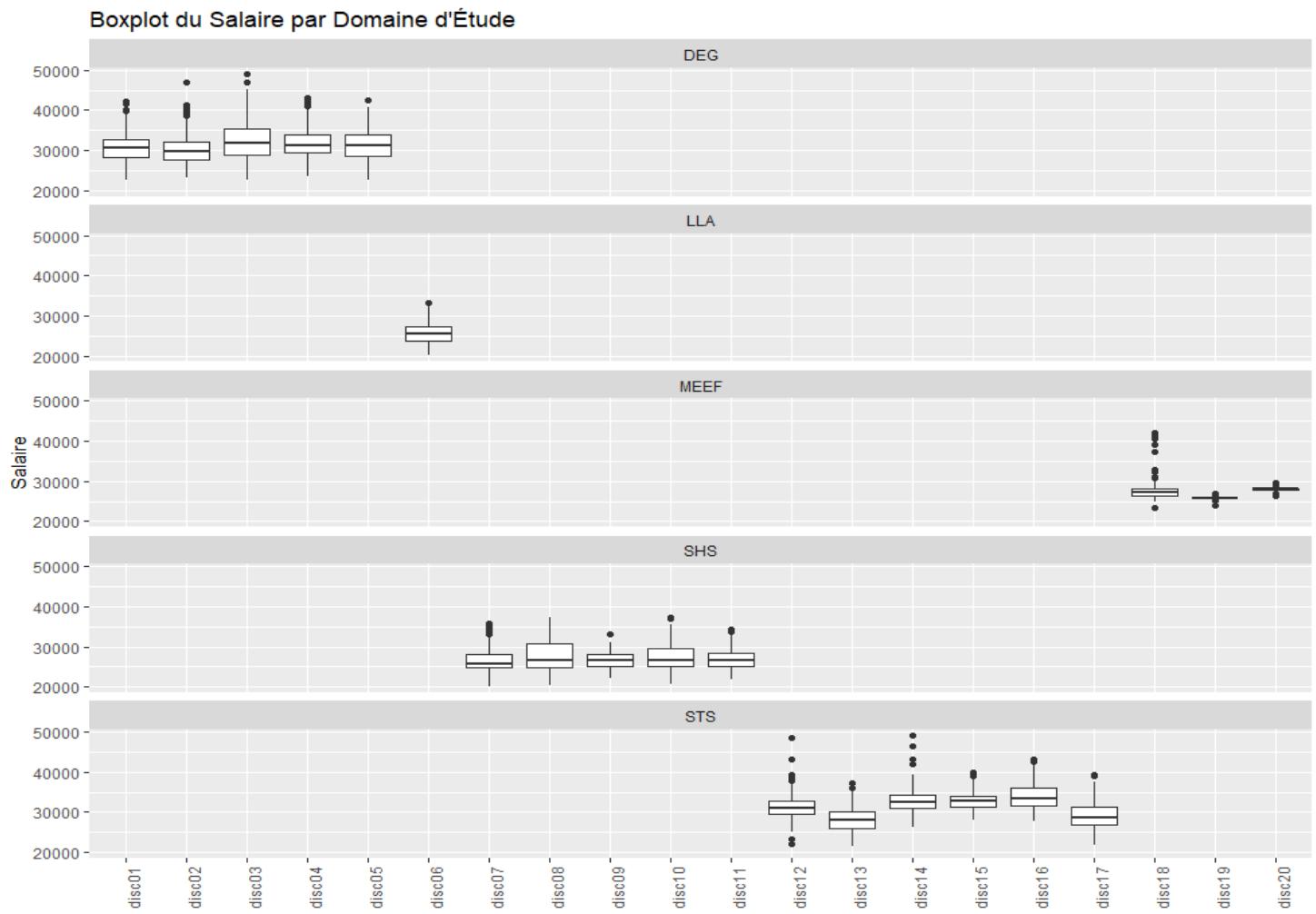
- Analyse des distributions du taux de chômage dans les différentes régions françaises :**



Nous constatons que la moyenne du taux du chômage dans presque toutes les régions de notre étude été entre 5% et 12%, sauf pour les régions du Guadeloupe et la réunion qui marque une moyenne très élevée, entre 17 et 19% entre 2010 et 2019.

- Analyse de la discipline avec le salaire brute :

Figure 8 : Distribution des salaires selon la spécialité.

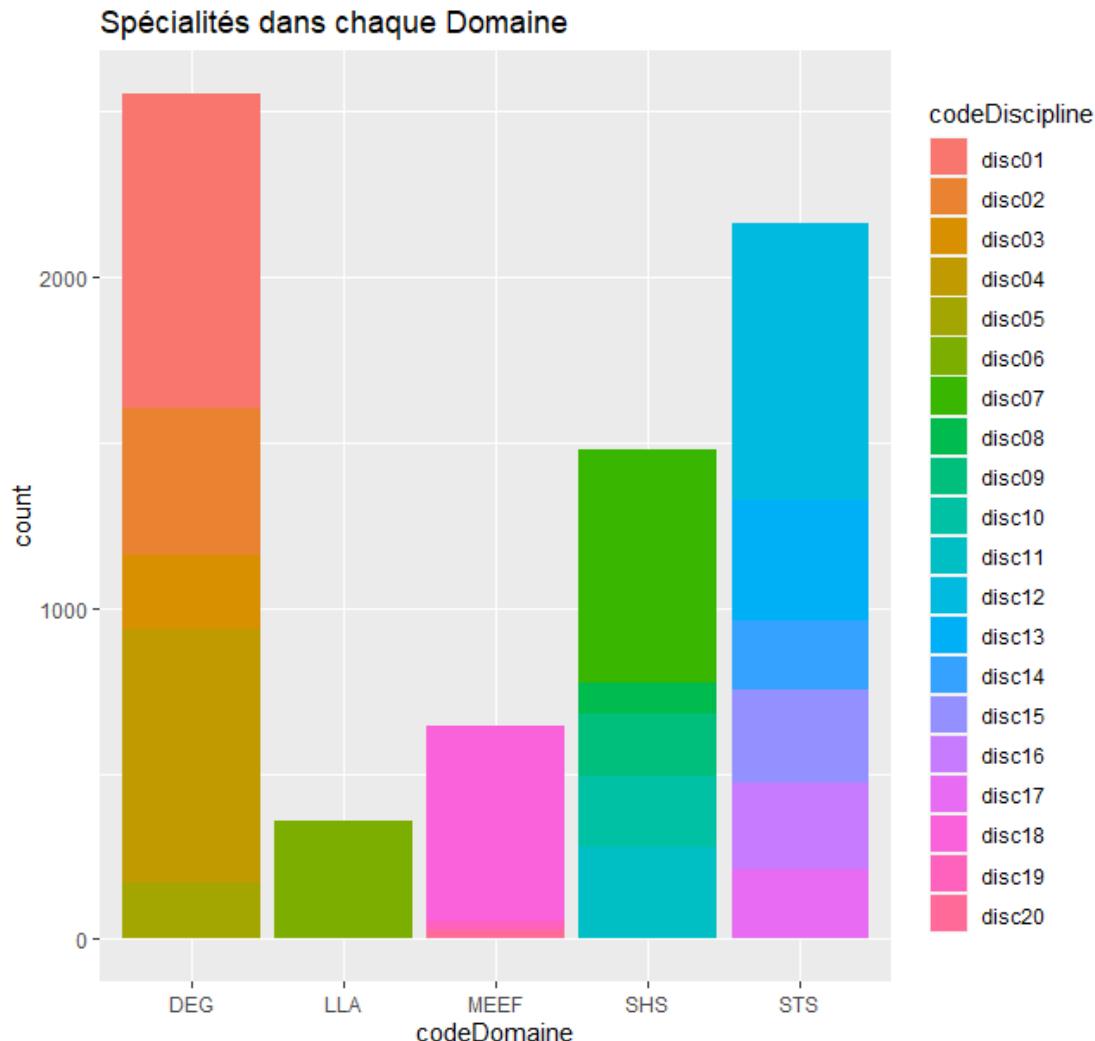


Nous constatons que le meilleur salaire brut se trouve dans les disciplines qui appartient aux domaines droit et gestion et Sciences technologiques et santé avec un salaire brute moyen entre 31000 et 35000 euros, avec un meilleur salaire pour la spécialité économie, Sciences Fondamentales, Sciences De L'ingénieur et Informatique.

D'autre part, nous remarquons que les disciplines qui ont le salaire brut le plus bas ont ceux qui appartient aux domaines lettre, langue et arts, Sciences humaine et sociales, master enseignement, avec un salaire brut entre 19000 et 25000 euros, avec la discipline du Lettres, Langues, Arts avec Ensemble Sciences Humaines Et Sociales et Masters Enseignement comme les disciplines avec le salaire brute le plus bas.

- Analyse des distributions du discipline avec le code du domaine :

Figure 9 : Distribution des disciplines selon le code du domaine.



Selon le graph, nous remarquons que les spécialités ensemble formations juridiques, économiques et de gestion avec la spécialité « droit » représente plus de 60 % de la totalité du domaine droit et gestion depuis 2010 au 2019

Pour le domaine sciences humaines et sociales nous remarquons que la spécialité histoire géographique est la plus dominante.

Nous constatons que les spécialités du domaine science technologique et santé sont équivalentes

Pour les domaines master enseignement et langue, lettre et arts les spécialités dominantes sont respectivement, langue, lettre et arts, avec master enseignement premier degré.

5. Construction des modèles :

5.1. Création du modèle de régression linéaire simple :

5.1.1. Etude de relation entre le taux d'insertion professionnel et poids de spécialité :

Le modèle que nous avons créé est une régression linéaire simple avec la variable **PoidsDiscipline** en tant que variable prédicteur et **tauxInsertion** comme variable dépendante.

Figure 10 : Résultat du premier modèle.

```

Console Terminal × Background Jobs ×
R 4.3.1 · ~/ ↗

Call:
lm(formula = tauxInsertion ~ PoidsDiscipline, data = my_data)

Residuals:
    Min      1Q  Median      3Q     Max 
-28.677 -3.303   0.795   3.979  11.187 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 88.628525   0.101934 869.47 <2e-16 ***
PoidsDiscipline 0.061345   0.002532  24.23 <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5.631 on 7195 degrees of freedom
Multiple R-squared:  0.07544, Adjusted R-squared:  0.07531 
F-statistic: 587.1 on 1 and 7195 DF,  p-value: < 2.2e-16

> |

```

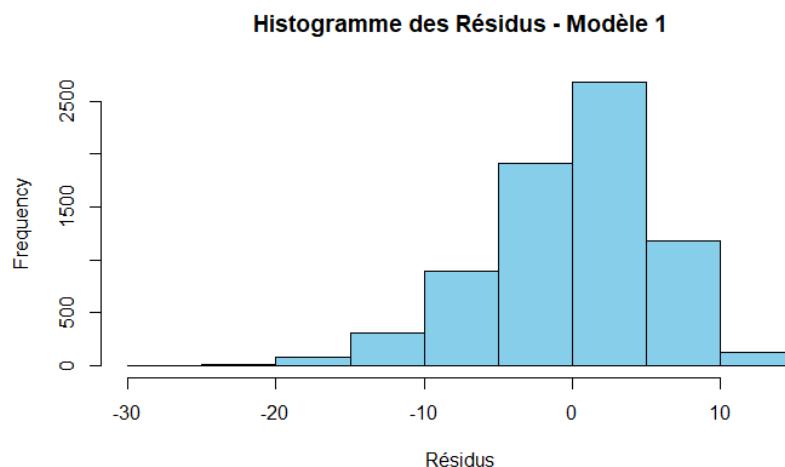
Le modèle de régression linéaire simple établi pour expliquer le taux d'insertion professionnelle en fonction du poids de la discipline révèle des résultats significatifs.

- Le coefficient positif de **PoidsDiscipline** (0.061345) indique que, en moyenne, pour chaque unité d'augmentation dans le poids de la discipline, le taux d'insertion professionnelle augmente d'environ 0.061345. Ce coefficient est statistiquement significatif avec un p-value très faible, suggérant que la variable **PoidsDiscipline** a un impact significatif sur le taux d'insertion.

- Cependant, le R-squared de 0.07544 indique que le modèle n'explique qu'environ 7.54% de la variation dans le taux d'insertion, soulignant la nécessité d'explorer d'autres variables qui pourraient influencer le taux d'insertion.

Visualisation de la distribution des résidus du modèle :

Figure 13 : Distribution des résidus.



La distribution des résidus nous montre que ces derniers ne suivent pas une loi normale.

5.1.2. Etude de relation entre le taux d'insertion professionnel et le taux de chômage :

Figure 14 : Résultat du deuxième modèle linéaire.

```

Console Terminal × Background Jobs ×
R 4.3.1 · ~/ 
Call:
lm(formula = tauxInsertion ~ tauxChomage, data = my_data)

Residuals:
    Min      1Q  Median      3Q     Max
-27.7146 -3.4640  0.7617  4.1915 11.1332

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 93.25131   0.35703 261.186 < 2e-16 ***
tauxChomage -0.31318   0.03992 -7.845 4.96e-15 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.832 on 7195 degrees of freedom
Multiple R-squared:  0.008481, Adjusted R-squared:  0.008343
F-statistic: 61.54 on 1 and 7195 DF,  p-value: 4.96e-15
> |

```

Interprétation des résultats :

1. Coefficients :

- Coefficient pour tauxChomage : Le coefficient est -0.31318. Cela signifie que, en moyenne, pour chaque augmentation d'une unité dans le tauxChomage, le taux d'insertion diminue d'environ 0.31318.

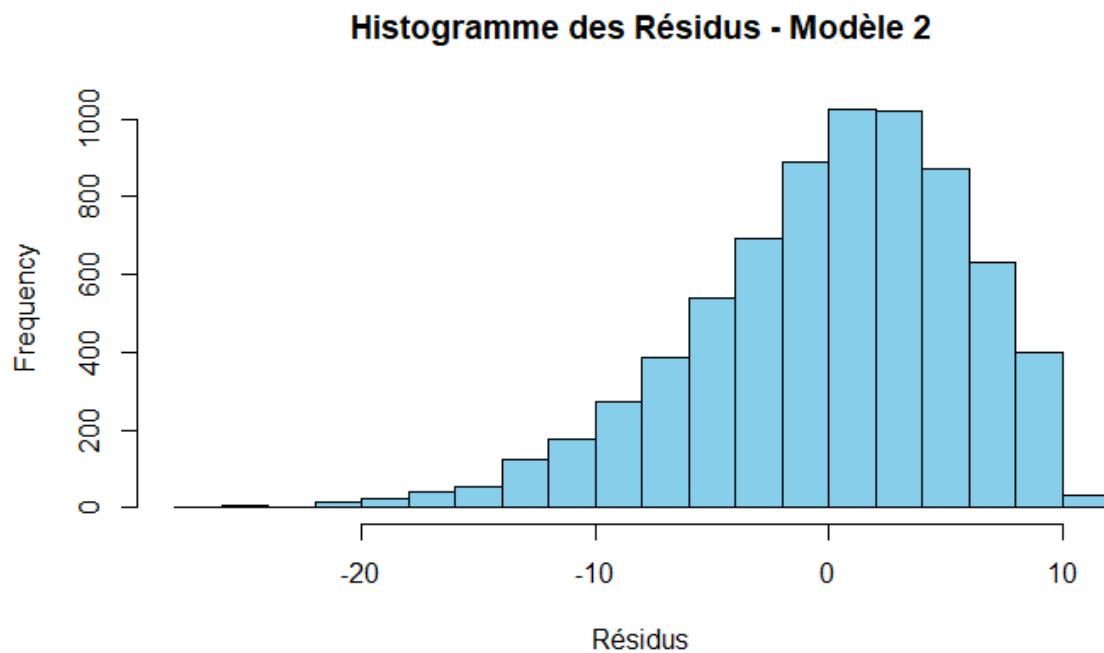
2. Test de Significativité :

- Le coefficient pour tauxChomage est significatif avec un p-value très faible, indiquant que le taux de chômage a un impact significatif sur le taux d'insertion.

3. R-squared et F-statistic :

- Le R-squared est de 0.008481, indiquant que seulement environ 0.85% de la variation dans le taux d'insertion est expliqué par le tauxChomage.
- L'Adjusted R-squared (corrigé pour le nombre de variables dans le modèle) est légèrement plus bas.

Figure 15 : Distribution des résidus.



La distribution des résidus nous montre que ces derniers ne suivent pas une loi normale.

On peut conclure que le modèle indique une relation significative entre le taux d'insertion professionnelle et le taux de chômage, la variation expliquée est limitée (R-squared de 0.008481).

5.1.3. Etude de relation entre le salaire brute et le taux de chômage :

Les résultats du modèle de régression linéaire simple entre le salaire brut (salaireBrute) et le taux de chômage (tauxChomage) sont les suivants :

Figure 16 : Résultat du modèle linéaire.

```

Console Terminal × Background Jobs ×
R 4.3.1 · ~/ ↗

Call:
lm(formula = my_data$salaireBrute ~ my_data$tauxChomage, data =
my_data)

Residuals:
    Min      1Q   Median      3Q     Max 
-9827.4 -2937.5  -352.5  2545.7 19010.9 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 33002.04   235.67 140.03 <2e-16 ***
my_data$tauxChomage -383.28    26.35 -14.54 <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3849 on 7195 degrees of freedom
Multiple R-squared:  0.02856, Adjusted R-squared:  0.02843 
F-statistic: 211.6 on 1 and 7195 DF, p-value: < 2.2e-16

> |

```

Interprétation des résultats :

1. Coefficients :

- Coefficient pour tauxChomage : Le coefficient est -383.28. Cela signifie que, en moyenne, pour chaque augmentation d'une unité dans le tauxChomage, le salaire brut diminue d'environ 383.28.

2. Test de Significativité :

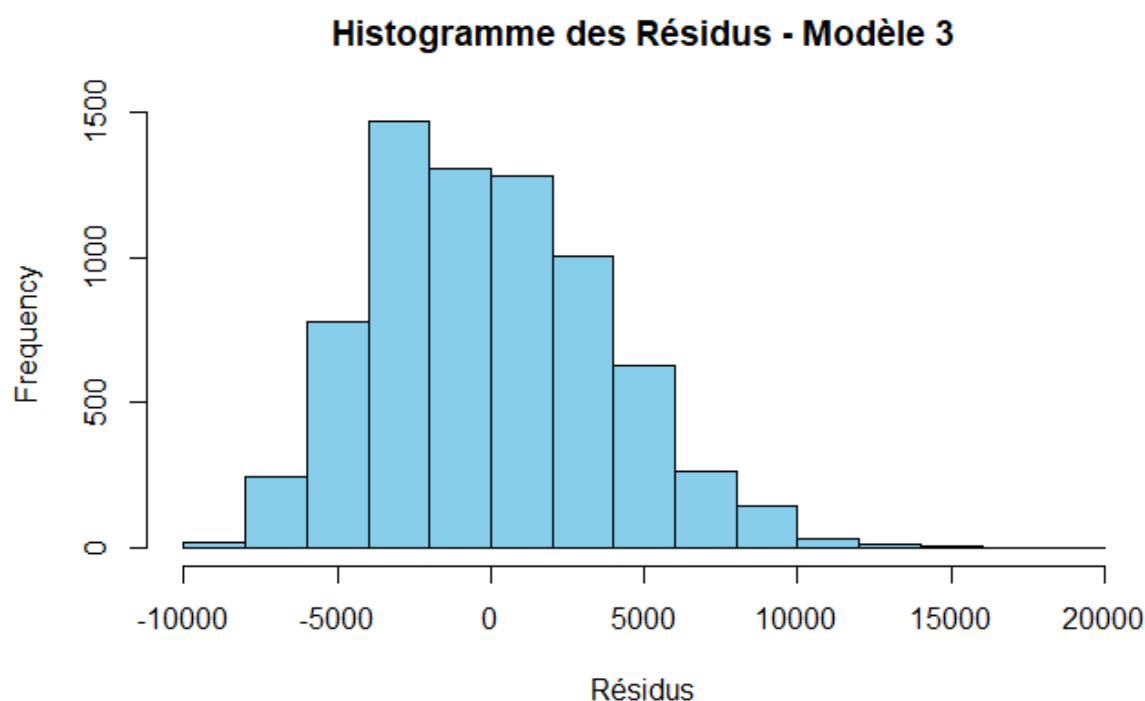
- Le coefficient pour tauxChomage est significatif avec un p-value très faible (<2e-16), indiquant que le taux de chômage a un impact significatif sur le salaire brut.

3. R-squared et F-statistic :

- Le R-squared est de 0.02856, indiquant que seulement environ 2.856% de la variation dans le salaire brut est expliqué par le tauxChomage.

Le modèle indique une relation significative entre le salaire brut et le taux de chômage, la variation expliquée est limitée (R-squared de 0.02856). D'autres facteurs non inclus dans ce modèle peuvent également jouer un rôle dans l'explication de la variation du salaire brut compréhension plus complète.

Figure 17 : Distribution des résidus.



Les résultats visuels nous montrent que les résidus de dernier modèle semblent suivre une loi normale.

5.2. Test d'homogénéité des modèles :

Les résultats du test de Breusch-Pagan indiquent la présence d'hétéroscédasticité dans les résidus des trois modèles.

Figure 18 : Résultat du test d'homogénéité des 3 modèles.

```

Console Terminal × Background Jobs ×
R 4.3.1 · ~/Documents
> bptest(lm(tauxInsertion ~ tauxChomage, data = my_data), studentize = FALSE)

Breusch-Pagan test

data: lm(tauxInsertion ~ tauxChomage, data = my_data)
BP = 11.024, df = 1, p-value = 0.0008995

> # Modèle 2 : Taux d'Insertion Professionnelle vs Code de Domaine Numérique
> bptest(lm(tauxInsertion ~ codeDomaine_numeric, data = my_data), studentize = FALSE)

Breusch-Pagan test

data: lm(tauxInsertion ~ codeDomaine_numeric, data = my_data)
BP = 104.34, df = 1, p-value < 2.2e-16

> # Modèle 3 : Salaire Brut vs Taux de Chômage
> bptest(lm(salaireBrute ~ tauxChomage, data = my_data), studentize = FALSE)

Breusch-Pagan test

data: lm(salaireBrute ~ tauxChomage, data = my_data)
BP = 20.9, df = 1, p-value = 4.839e-06

```

Modèle 1 : Taux d'Insertion Professionnelle vs Taux de Chômage

- Statistique du test (BP) : 11.024
- Valeur p (p-value) : 0.0008995

La valeur p de 0.0008995 est inférieure à un niveau de signification communément choisi comme 0.05. Cela suggère que l'hypothèse nulle d'homoscédasticité est rejetée, indiquant la présence d'hétéroscédasticité dans les résidus.

Modèle 2 : Taux d'Insertion Professionnelle vs Code de Domaine Numérique

- Statistique du test (BP) : 104.34
- Valeur p (p-value) : < 2.2e-16

La valeur p très faible (< 2.2e-16) indique également le rejet de l'hypothèse d'homoscédasticité dans ce modèle.

Modèle 3 : Salaire Brut vs Taux de Chômage

- Statistique du test (BP) : 20.9
- Valeur p (p-value) : 4.839e-06

La valeur p très faible (< 0.000001) indique que l'hypothèse d'homoscédasticité est rejetée pour ce modèle.

En conclusion, ces résultats suggèrent que l'hétéroscédasticité est présente dans les résidus des modèles.

5.3. Crédit au modèle de régression linéaire multiple :

5.3.1. Salaire brute en fonction du taux de chômage et taux d'insertion et poids de discipline :

Le modèle que nous souhaitons créer vise à expliquer le salaire brut en fonction du taux d'insertion professionnelle et du taux de chômage. En examinant l'impact de ces deux variables sur le salaire, nous cherchons à comprendre comment les opportunités d'emploi et le contexte économique influencent directement la rémunération des individus. Cette modélisation nous permettra de tirer des conclusions essentielles sur les dynamiques salariales en relation avec le marché du travail et le chômage.

Figure 18 : Résumé statistique du modèle.

```

Console Terminal × Background Jobs ×
R 4.3.1 · ~/ ↵

Call:
lm(formula = salaireBrute ~ tauxInsertion + tauxChomage + PoidsDiscipline,
  data = my_data)

Residuals:
    Min      1Q   Median      3Q     Max 
-10175.9 -2575.3  -309.3   2214.9  18407.3 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 8318.742    714.342   11.64 <2e-16 ***
tauxInsertion 272.390     7.419   36.71 <2e-16 ***
tauxChomage -285.535    24.272  -11.76 <2e-16 ***
PoidsDiscipline -27.053     1.651  -16.39 <2e-16 ***  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3525 on 7193 degrees of freedom
Multiple R-squared:  0.1858,    Adjusted R-squared:  0.1854 
F-statistic: 547.1 on 3 and 7193 DF,  p-value: < 2.2e-16

> |

```

- Taux d'insertion (tauxInsertion) : Le coefficient associé au taux d'insertion est d'environ 0.358. Cela signifie qu'une augmentation d'une unité du taux d'insertion professionnel est associée à une augmentation de 0.358 unité du salaire brut, Ce coefficient est statistiquement significatif (***) , ce qui indique une relation positive entre le taux d'insertion et le salaire brut.
- Taux de chômage (tauxChomage) : Le coefficient associé au taux de chômage est d'environ -0.136. Cela signifie qu'une augmentation d'une unité du taux de chômage est associée à une diminution de 0.136 unité du salaire brut, Ce coefficient est également statistiquement significatif (***) , ce qui indique une relation négative entre le taux de chômage et le salaire brut.
- PoidsDiscipline: Le coefficient associé au taux de chômage est d'environ -0.136. Cela signifie qu'une unité d'augmentation dans PoidsDiscipline, le salaire brut diminue en moyenne de 27.053 dollars.

Les coefficients sont interprétés en supposant que toutes les autres variables restent constantes. Le modèle a une faible capacité explicative avec un R^2 de 0.1854, ce qui signifie que seulement 18.54 % de la variance du salaire brut est expliquée par les variables du modèle. Le modèle a une grande significativité globale avec un F-statistic élevé (661.7) et un p-value très faible (< 2.2e-16), indiquant que l'ensemble du modèle est statistiquement significatif.

5.3.2. Etude de salaire brut en fonction de la spécialité :

Pour étudier le salaire brut en fonction de la spécialité d'étude, nous devons suivre plusieurs étapes, notamment la création d'un modèle linéaire multiple avec des variables indicatrices pour chaque spécialité.

- a) Création de la variable indicatrices « **Dummies** » qui regroupe les différentes modalités de la variable codeDiscipline.
- b) Utiliser la fonction « log » pour la variable salaire brute afin de donner plus de linéarité au modèle.
- c) Création de modèle.

Figure 19 : Résultat de modèle

```

Console Terminal Background Jobs
R 4.3.1 - ~/R

coefficients: (1 not defined because of singularities)
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 28126.9231   603.8329  46.581 < 2e-16 ***
dummiescodeDisciplinedisc01 2635.2220   612.0315  4.306 1.69e-05 ***
dummiescodeDisciplinedisc02 2008.4373   621.2611  3.233 0.00123 **
dummiescodeDisciplinedisc03 4262.4575   637.6213  6.685 2.48e-11 ***
dummiescodeDisciplinedisc04 3694.8785   613.9952  6.018 1.86e-09 ***
dummiescodeDisciplinedisc05 2964.1484   648.8779  4.568 5.00e-06 ***
dummiescodeDisciplinedisc06 -2542.2862   625.3755 -4.065 4.85e-05 ***
dummiescodeDisciplinedisc07 -1758.2262   614.8511 -2.860 0.00425 **
dummiescodeDisciplinedisc08 -472.5753    683.8547 -0.691 0.48956
dummiescodeDisciplinedisc09 -1534.3305   644.0285 -2.382 0.01723 *
dummiescodeDisciplinedisc10 -737.3004    639.7898 -1.152 0.24919
dummiescodeDisciplinedisc11 -1257.0306   631.3418 -1.991 0.04651 *
dummiescodeDisciplinedisc12 3041.9392   613.1618  4.961 7.17e-07 ***
dummiescodeDisciplinedisc13 0.2508     624.7999  0.000 0.99968
dummiescodeDisciplinedisc14 4550.1104   640.2912  7.106 1.31e-12 ***
dummiescodeDisciplinedisc15 4717.0055   631.2457  7.473 8.80e-14 ***
dummiescodeDisciplinedisc16 5928.8909   633.5284  9.359 < 2e-16 ***
dummiescodeDisciplinedisc17 1412.5135    639.6257  2.208 0.02725 *
dummiescodeDisciplinedisc18 -672.2541    617.0163 -1.090 0.27596
dummiescodeDisciplinedisc19 -2355.4945   838.5609 -2.809 0.00498 **
dummiescodeDisciplinedisc20 NA          NA        NA        NA
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3079 on 7177 degrees of freedom
Multiple R-squared:  0.3801, Adjusted R-squared:  0.3784
F-statistic: 231.6 on 19 and 7177 DF,  p-value: < 2.2e-16
> |

```

En se basant sur ces résultats :

- **Significatifs (p-value < 0.05) :**

Les spécialités des domaines droit et gestion, sciences humaines et sociales, sciences techniques et informatiques sont tous significatifs au seuil de 5%.

Nous remarquons aussi que les spécialités du domaine informatique et sciences d'ingénieur plus le domaine du droit et gestion ont la plus d'impact dans le salaire brut, en regardant par exemple l'augmentation d'une unité de la spécialité sciences de l'ingénieur augmentera le salaire avec plus de 18,71.

- **Non significatifs (p-value >= 0.05) :**

Nous pouvons voir aussi que les spécialités histoires et géographie, communication, master et enseignement n'ont pas significatifs au seuil de 5% pour la variable salaire brut.

- **NA (disc20) :**

La modalité master et enseignement second degré n'est pas renseigné en raison de singularités, Le problème de singularité se produit dans le contexte de l'analyse de régression lorsque certaines variables explicatives dans le modèle sont linéairement dépendantes les unes des autres. En d'autres termes, une variable peut être exprimée comme une combinaison linéaire des autres, ce qui crée une instabilité dans l'estimation des coefficients du modèle, notre modalité master second degré et liée au modalité master premier degré.

Nous pouvons voir que le R-carré égale à 37,84% de la qualité d'ajustement de la variable salaire brute, on peut conclure que les spécialités d'étude expliquent une partie importante des salaires brutes en France.

5.3.3. Création de modèle multiple globale :

Nous allons construire un modèle linéaire multiple avec plusieurs variables explicatives, y compris des variables indicatrices pour la discipline (dummies), le tauxInsertion, et le tauxChomage pour prédire la variable dépendante salaireBrute.

Figure 20 : Résultat de modèle globale.

```
Console Terminal × Background Jobs ×
R 4.3.1 · ~/ ↗

Coefficients: (1 not defined because of singularities)
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    11997.144   932.484 12.866 < 2e-16 ***
dummiescodeDisciplinedisc01 3556.207   566.452  6.278 3.63e-10 ***
dummiescodeDisciplinedisc02 2728.990   574.778  4.748 2.10e-06 ***
dummiescodeDisciplinedisc03 5290.970   591.195  8.950 < 2e-16 ***
dummiescodeDisciplinedisc04 4262.942   567.572  7.511 6.59e-14 ***
dummiescodeDisciplinedisc05 4026.345   601.017  6.699 2.25e-11 ***
dummiescodeDisciplinedisc06 -543.571   583.911 -0.931 0.351930
dummiescodeDisciplinedisc07 -19.621   572.889 -0.034 0.972680
dummiescodeDisciplinedisc08 1637.681   637.732  2.568 0.010249 *
dummiescodeDisciplinedisc09 -669.569   595.858 -1.124 0.261175
dummiescodeDisciplinedisc10 794.420   593.940  1.338 0.181088
dummiescodeDisciplinedisc11 532.249   588.648  0.904 0.365925
dummiescodeDisciplinedisc12 4074.818   568.092  7.173 8.08e-13 ***
dummiescodeDisciplinedisc13 1973.853   583.204  3.385 0.000717 ***
dummiescodeDisciplinedisc14 5823.386   593.478  9.812 < 2e-16 ***
dummiescodeDisciplinedisc15 5421.587   583.450  9.292 < 2e-16 ***
dummiescodeDisciplinedisc16 5857.361   584.436 10.022 < 2e-16 ***
dummiescodeDisciplinedisc17 2233.138   591.422  3.776 0.000161 ***
dummiescodeDisciplinedisc18 -1091.348   569.163 -1.917 0.055219 .
dummiescodeDisciplinedisc19 -2227.675   773.058 -2.882 0.003968 **
dummiescodeDisciplinedisc20      NA        NA        NA
tauxInsertion       200.849    7.138 28.137 < 2e-16 ***
tauxChomage        -346.235   19.718 -17.560 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2838 on 7175 degrees of freedom
Multiple R-squared:  0.4734,    Adjusted R-squared:  0.4718
F-statistic: 307.1 on 21 and 7175 DF,  p-value: < 2.2e-16
```

Nous pouvons conclure qu'il existe plusieurs changements de coefficient au niveau des modalités (la spécialité informatique n'est plus significative au niveau de 5%).

Les autres modalités, le taux de d'insertion et le taux de chômage sont significatives (comme c'est montré dans les précédents modèles).

Nous procéderons avec ce modèle grâce au R-carré qui est égal à 0.4718, cela veut dire que les variables de notre modèle expliquent 47.18% de notre variable dépendante.

On doit tout d'abord exclure les modalités qui ne sont pas significatives pour donner plus de linéarité à notre modèle (**voir annexe 01**)

6. Performance du modèle :

Figure 12 : Test de performance.

```
```{r}
library(performance)
check_collinearity(model_eco_3)
```

# Check for Multicollinearity

Low Correlation



Term	VIF	VIF 95% CI	Increased SE
dummies[, -c(6, 7, 9, 10, 11, 18, 19, 20)]	1.13	[1.11, 1.17]	1.07
tauxInsertion	1.14	[1.11, 1.17]	1.07
tauxChomage	1.02	[1.00, 1.07]	1.01



Tolerance Tolerance 95% CI
0.88 [0.86, 0.90]
0.88 [0.85, 0.90]
0.98 [0.93, 1.00]
```

Selon le test de la performance de modèle on a le VIF (facteur d'inflation de la variance) est inférieur à 5% donc le modèle qu'on a obtenu n'est pas affecté par un problème de multicolonéarité.

- **Low Correlation :**

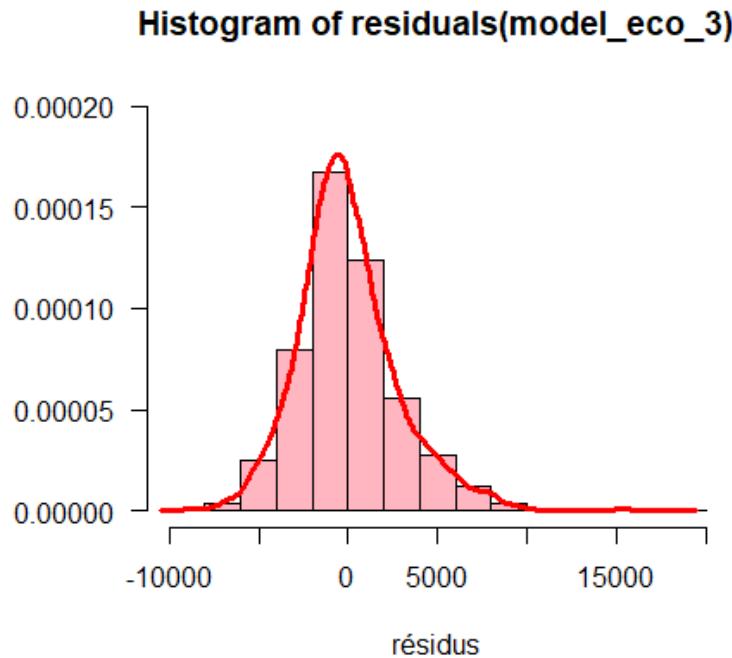
- Les variables incluses dans la première ligne (dummies) ont un VIF de 1.13, ce qui indique une faible corrélation avec les autres variables du modèle. Un VIF proche de 1 suggère une faible multicollinéarité.
- Les variables "tauxInsertion" et "tauxChomage" ont également des VIF assez bas (1.14 et 1.02 respectivement), ce qui suggère une faible corrélation avec les autres variables.

- **Tolerance :**

- La tolérance est l'inverse du VIF ($\text{Tolerance} = 1/\text{VIF}$). Une tolérance proche de 1 indique une faible multicollinéarité. Dans notre cas, toutes les tolérances sont proches de 1, ce qui est généralement considéré comme bon.

- Normalité des résidus :

Figure 13 : Distribution des résidus



Test Anderson-Darling (normalité) :

```
```{r}
#Anderson-Darling normality test
library(nortest)
ad.test(model_eco_3$residuals)
```

Anderson-Darling normality test

```
data: model_eco_3$residuals
A = 51.25, p-value < 2.2e-16
```

Les résultats du test d'Anderson-Darling indiquent une forte évidence contre l'hypothèse nulle que les résidus suivent une distribution normale.

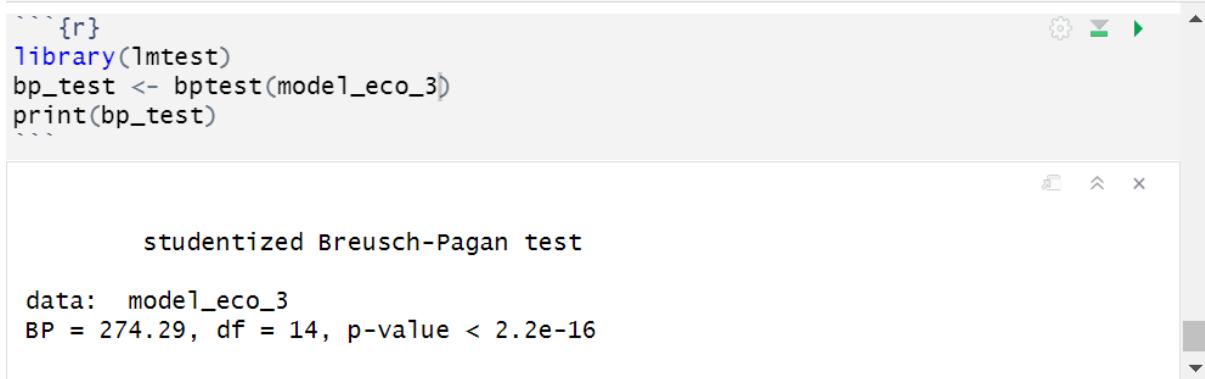
- Statistique de test (A) : La statistique de test (A) est de 51.25. Cette valeur est utilisée pour évaluer la déviation des résidus par rapport à une distribution normale. Plus la statistique A est élevée, plus il y a de preuves contre la normalité.
- P-valeur : La p-valeur est très petite, Cela signifie que la p-valeur est essentiellement nulle. Il y a une très faible probabilité d'observer ces données si les résidus suivaient une distribution normale.
- Interprétation :

Comme nous avons vu, la p-valeur est extrêmement faible (inférieure au seuil de 0.05), on rejette l'hypothèse nulle. Ainsi, les résidus ne suivent pas une distribution normale selon le test d'Anderson-Darling.

Lorsque la normalité des résidus est violée, il peut être utile d'explorer d'autres transformations de données, d'identifier et de traiter les valeurs aberrantes dans notre cas on ne dispose pas des valeurs aberrantes, ou de considérer des modèles statistiques plus complexes qui peuvent mieux s'adapter aux données.

En tout cas, il est important d'investiguer davantage la nature des résidus pour comprendre les raisons de la non-normalité.

- **Test d'homoscédasticité :**



```
```{r}
library(lmtest)
bp_test <- bptest(model_eco_3)
print(bp_test)
```

studentized Breusch-Pagan test

data: model_eco_3
BP = 274.29, df = 14, p-value < 2.2e-16

La statistique de test BP est de 274.29 tandis que le test a été effectué avec 14 degrés de liberté.
La p-valeur associée au test est pratiquement nulle, étant inférieure à 2.2e-16.

Ces résultats suggèrent fortement que les résidus ne suivent pas une distribution homoscédastique.

En d'autres termes, il y a des preuves statistiques solides d'une hétéroscédasticité significative, indiquant que la variance des résidus varie considérablement à travers les niveaux des variables explicatives.

Conclusion du modèle :

Ces constatations soulignent la possibilité que l'hypothèse d'homoscédasticité n'est pas satisfaite dans notre modèle économétrique. Lorsque l'homoscédasticité est enfreinte, des ajustements sont souvent nécessaires pour tenir compte de la variance inégale des résidus, tels que la transformation des données, l'introduction de poids dans la régression, ou l'utilisation d'autres méthodes de modélisation adaptées à l'hétéroscléasticité.

7. Vérification des hypothèses :

- **Hypothèse 1 : (Acceptée)** Il est courant de supposer que le niveau de diplôme est étroitement lié au salaire brut annuel estimé. En général, les diplômés de niveaux supérieurs, tels que les titulaires de diplômes universitaires avancés, ont tendance à gagner en moyenne des salaires plus élevés que ceux ayant des diplômes de niveau inférieur. Cette hypothèse est souvent confirmée par des données empiriques et est cohérente avec l'idée que l'éducation supérieure et la formation avancée peuvent augmenter la valeur sur le marché du travail.
- **Hypothèse 2 : (Acceptée)** L'idée que le taux de chômage national puisse influencer les opportunités d'emploi est valide. En périodes de chômage élevé, la concurrence sur le marché du travail devient plus féroce, ce qui peut réduire les perspectives d'emploi pour les individus. Les employeurs peuvent avoir plus de choix parmi les candidats, ce qui peut exercer une pression à la baisse sur les salaires. Les fluctuations du taux de chômage ont un impact sur l'offre et la demande de main-d'œuvre, et donc sur les salaires.
- **Hypothèse 3 : (Acceptée)** L'idée que le choix du domaine d'études puisse avoir un impact sur les revenus est valide. Certains domaines d'études, tels que les sciences, la technologie, l'ingénierie et les mathématiques, gestion et économie, ont tendance à offrir des opportunités plus lucratives sur le marché du travail en raison de la demande élevée de compétences spécialisées.

8. Recommandations :

Les constatations de cette analyse suggèrent des disparités significatives dans les salaires et le taux de chômage entre les régions et les spécialités. Pour une meilleure équité salariale et des opportunités d'emploi égales, il pourrait être utile de mettre en place des politiques de réduction des inégalités régionales et sectorielles, en se concentrant sur les régions présentant des taux de chômage élevés.

De plus, les étudiants et les professionnels cherchant à maximiser leur potentiel de rémunération peuvent envisager de se tourner vers des spécialités relevant des domaines du droit, de la gestion, des sciences fondamentales, de l'ingénierie et de l'informatique, qui offrent des salaires bruts plus élevés en moyenne.

Les établissements d'enseignement et les décideurs politiques peuvent tirer des enseignements de cette analyse pour orienter leurs programmes et ressources en fonction des besoins régionaux et sectoriels. L'éducation et la formation dans les domaines à fort potentiel économique pourraient aider à réduire les disparités salariales.

En résumé, l'analyse souligne l'importance de la prise en compte des facteurs régionaux et sectoriels dans les choix d'éducation et de carrière, tout en appelant à des politiques visant à réduire les inégalités et à favoriser la croissance économique équilibrée.

9. Conclusion

Notre analyse met en évidence des variations significatives des salaires bruts en France, avec des disparités régionales notables. Les régions du sud, de l'est, et de l'ouest, telles que Lyon, Normandie, et Orléans, présentent des salaires bruts moyens proches de 26 000 euros. En revanche, les régions parisiennes, Île-de-France, Guadeloupe et La Réunion se démarquent avec des salaires bruts moyens atteignant environ 35 000 euros. L'examen du taux de chômage révèle que la plupart des régions ont maintenu un taux compris entre 5% et 12%, exception faite de Guadeloupe et La Réunion, qui ont enregistré des taux de chômage nettement plus élevés, oscillant entre 17% et 19% sur la période 2010-2019.

En ce qui concerne les spécialités, les domaines du droit, de la gestion, des sciences fondamentales, des sciences de l'ingénieur et de l'informatique ont les salaires bruts les plus élevés, variant de 31 000 à 35 000 euros en moyenne. Plus précisément, la spécialité économie semble offrir le salaire brut le plus attractif. À l'inverse, les spécialités relevant des domaines des lettres, des langues, des arts, des sciences humaines et sociales, ainsi que le domaine de l'enseignement, affichent des salaires bruts plus modestes, allant de 19 000 à 25 000 euros. Parmi celles-ci, les spécialités "Lettres, Langues, Arts" et "Ensemble Sciences Humaines Et Sociales" enregistrent les salaires bruts les plus bas.

De plus, une observation du graphique révèle que les spécialités "Ensemble Formations Juridiques, Économiques et de Gestion" avec la spécialité "Droit" dominent largement, représentant plus de 60% de l'ensemble du domaine Droit et Gestion de 2010 à 2019. Dans le domaine des sciences humaines et sociales, la spécialité "Histoire-Géographie" se démarque comme la plus prédominante. Pour le domaine des sciences technologiques et de la santé, les spécialités sont relativement équivalentes. Enfin, pour les domaines "Master Enseignement" et "Langue, Lettre et Arts", les spécialités "Langue, Lettre et Arts" et "Master Enseignement Premier Degré" dominent respectivement.

En outre, les tests statistiques confirment nos suppositions initiales :

- Le niveau de diplôme est étroitement lié aux salaires bruts, avec des diplômés de niveaux supérieurs gagnant en moyenne des salaires plus élevés.

- Le taux de chômage national a une influence significative sur les opportunités d'emploi, avec un taux de chômage plus élevé réduisant les perspectives d'emploi pour les individus.
- Les tests mettent en évidence des écarts salariaux entre les sexes, même après avoir contrôlé d'autres facteurs.
- Le choix du domaine d'études a un impact sur les revenus, avec certains domaines plus lucratifs que d'autres.
- Le niveau de diplôme influe sur le taux d'insertion des diplômés dans leur domaine d'études, avec de meilleures perspectives d'insertion pour les diplômés de niveau supérieur.

Ces constatations soulignent l'importance de prendre en compte les facteurs régionaux, sectoriels et éducatifs dans les choix de carrière et d'éducation. Les inégalités salariales et les disparités entre les sexes sont des préoccupations cruciales qui nécessitent une attention continue. Les résultats des tests renforcent nos conclusions précédentes concernant les disparités salariales, les inégalités entre les sexes et les choix éducatifs et professionnels.

10. Annexe

Annexe 01 :

```

Console Terminal × Background Jobs ×
R 4.3.1 · ~/Documents/R/Projet_Statistique/Chapitre_10/Exercice_1.R
Call:
lm(formula = salaireBrute ~ dummies[, -c(6, 7, 9, 10, 11, 18,
19, 20)] + tauxInsertion + tauxChomage, data = my_data)

Residuals:
    Min      1Q  Median      3Q     Max 
-9397.6 -1767.1  -297.2  1433.6 18383.1 

Coefficients:
                Estimate Std. Error t value Pr(>|t|)    
(Intercept) 14087.963   598.047  23.557 < 2e-16 ***
dummies[, -c(6, 7, 9, 10, 11, 18, 19, 20)]codeDisciplinedisc01 3895.296   110.001  35.411 < 2e-16 ***
dummies[, -c(6, 7, 9, 10, 11, 18, 19, 20)]codeDisciplinedisc02 3081.257   148.216  20.789 < 2e-16 ***
dummies[, -c(6, 7, 9, 10, 11, 18, 19, 20)]codeDisciplinedisc03 5592.157   199.000  28.101 < 2e-16 ***
dummies[, -c(6, 7, 9, 10, 11, 18, 19, 20)]codeDisciplinedisc04 4639.199   120.006  38.658 < 2e-16 ***
dummies[, -c(6, 7, 9, 10, 11, 18, 19, 20)]codeDisciplinedisc05 4338.725   228.174  19.015 < 2e-16 ***
dummies[, -c(6, 7, 9, 10, 11, 18, 19, 20)]codeDisciplinedisc08 1810.836   305.310   5.931 3.15e-09 ***
dummies[, -c(6, 7, 9, 10, 11, 18, 19, 20)]codeDisciplinedisc12 4392.739   114.987  38.202 < 2e-16 ***
dummies[, -c(6, 7, 9, 10, 11, 18, 19, 20)]codeDisciplinedisc13 2164.660   162.349  13.333 < 2e-16 ***
dummies[, -c(6, 7, 9, 10, 11, 18, 19, 20)]codeDisciplinedisc14 6119.185   206.245  29.669 < 2e-16 ***
dummies[, -c(6, 7, 9, 10, 11, 18, 19, 20)]codeDisciplinedisc15 5790.677   181.306  31.939 < 2e-16 ***
dummies[, -c(6, 7, 9, 10, 11, 18, 19, 20)]codeDisciplinedisc16 6320.736   191.279  33.045 < 2e-16 ***
dummies[, -c(6, 7, 9, 10, 11, 18, 19, 20)]codeDisciplinedisc17 2588.103   204.833  12.635 < 2e-16 ***
tauxInsertion                               175.377    6.144   28.547 < 2e-16 ***
tauxChomage                                 -358.573   19.720  -18.183 < 2e-16 ***

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

Residual standard error: 2858 on 7182 degrees of freedom
Multiple R-squared:  0.4653,    Adjusted R-squared:  0.4643 
F-statistic: 446.4 on 14 and 7182 DF,  p-value: < 2.2e-16

```

11. Bibliographie

- 1- Autor, D. H., Katz, L. F., & Kearney, M. S. (2008). Trends in U.S. Wage Inequality: Revising the Revisionists. *The Review of Economics and Statistics*, 90(2), 300-323.
- 2- Blau, F. D., & Kahn, L. M. (2017). The gender wage gap: Extent, trends, and explanations. *Journal of Economic Literature*, 55(3), 789-865.
- 3- Goldin, C. (2014). A grand gender convergence: Its last chapter. *American Economic Review*, 104(4), 1091-1119.
- 4- Acemoglu, D., & Autor, D. (2011). Skills, tasks and technologies: Implications for employment and earnings. *Handbook of Labor Economics*, 4, 1043-1171.
- 5- Card, D., & Krueger, A. B. (1994). Minimum wages and employment: A case study of the fast-food industry in New Jersey and Pennsylvania. *The American Economic Review*, 84(4), 772-793.
- 6- Becker, G. S. (1964). *Human capital: A theoretical and empirical analysis, with special reference to education*. Columbia University Press.
- 7- Blundell, R., Dearden, L., Goodman, A., & Reed, H. (2000). The returns to higher education in Britain: Evidence from a British cohort. *The Economic Journal*, 110(461), F82-F99.

12. Source

- 1- La base de données de notre étude : <https://www.data.gouv.fr/fr/datasets/insertion-professionnelle-des-diplomes-de-master-en-universites-et-etablissements-assimiles/>
- 2- Validation de la base de données par le ministère de l'enseignement supérieur :
https://data.enseignementsup-recherche.gouv.fr/explore/dataset/fr-esr-insertion_professionnelle-master/information/?disjunctive.numero_de_l_etablissement&disjunctive.academie