



ENSAE PARISTECH

PROJET DE MICRO-ÉCONOMÉTRIE APPLIQUÉE

---

# **L'effet de l'expérience sur l'emploi chez les jeunes**

---

MARIANNE SORBA  
FRANÇOIS TRAINAR

*Encadrant :*  
YANNICK GUYONVARCH

19 Mai 2017

# 1 Introduction

La crise bancaire et financière de 2008 a eu un effet considérable sur le marché du travail. Les pays faisant partie de l'Union européenne ont tous été confrontés à une montée importante du chômage. Cette hausse a touché par dessus-tout les jeunes entre 15 et 29 ans. Malgré les différentes mesures qui ont été mises en place pour soutenir l'emploi depuis la crise, le taux de chômeurs reste aujourd'hui extrêmement élevé avec un taux de 9.6% en France et de 24.6% chez les jeunes en 2016. Pourtant, le niveau d'étude n'a fait que croître dans le temps, et les jeunes sont plus souvent diplômés de l'enseignement supérieur que les personnes plus âgées ( 42 % des jeunes diplômés de l'enseignement supérieur). C'est aux vues de ces chiffres et de l'importance du sujet que dans le cadre de ce projet, nous avons décidé de nous intéresser au dynamique de l'emploi et du chômage chez les jeunes. Nous avons choisi comme problématique **l'effet de l'expérience sur l'emploi chez les jeunes**. Notre objectif fut d'essayer de comprendre comment agit l'expérience (académique comme professionnelle comme les stages, diplômes, etc...) sur les chances de trouver un emploi chez les jeunes.

## 2 Statistiques Descriptives de la base de données

### 2.1 Description de la base et des variables

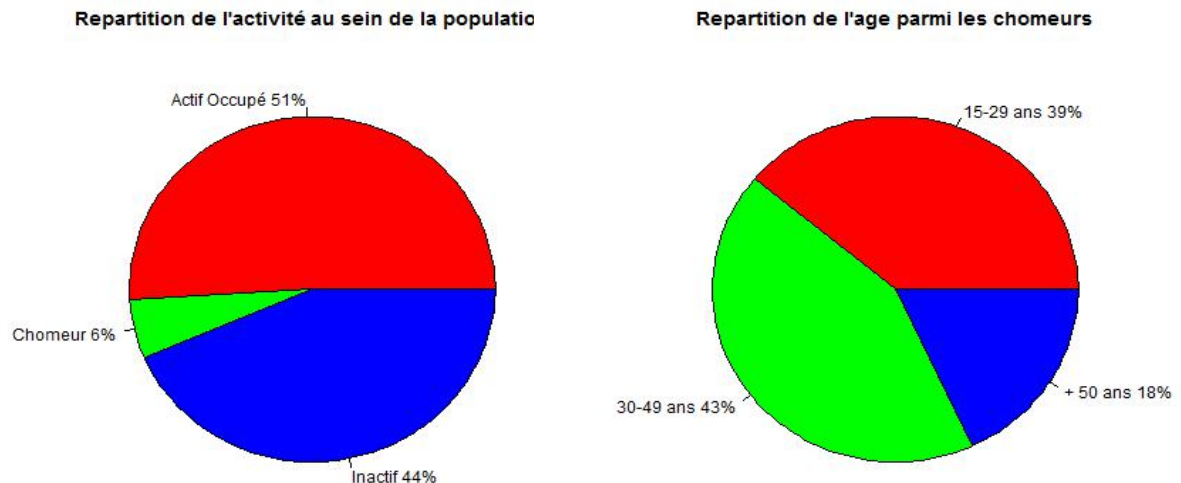
Pour mener notre étude, nous avons utilisé la base de données "Enquête Emploi en continu de 2013". Cette enquête, réalisée par L'Institut National de la Statistique et des Études Économiques (INSEE) en 2013, vise à observer à la fois de manière structurelle et conjoncturelle la situation des personnes sur le marché du travail en France. Chaque individu faisant parti de l'enquête a dû répondre à un questionnaire portant sur son activité professionnelle. La base de données contient exclusivement des variables qualitatives, avec un total de 122 variables et 394 294 individus (France, métropole et DOM). Chaque variable correspond à la réponse à une des questions du questionnaire. Les variables portent sur l'emploi, le chômage (défini en Annexe 1), la formation, l'origine sociale, la situation professionnelle un an auparavant, et la situation principale mensuelle sur les douze derniers mois.

Afin d'obtenir une base de données exploitable, nous avons en premier lieu procédé à un tri de la base de données de l'INSEE. Notre base de données finale contient des variables qui sont exclusivement reliées au profil de l'individu, à son expérience, à son niveau d'activité professionnelle et à son temps de chômage (s'il est effectivement chômeur). Nous avons aussi supprimé les inactifs (étudiant, homme/femme au foyer, retraités, etc.) de notre base de données parce qu'ils ne nous intéressent pas dans le cadre de notre étude. Vous trouverez la liste des variables de notre base de données final en Annexe 2.

### 2.2 Justification de la problématique

Avant de rentrer dans le vif du sujet, il convient de justifier notre problématique : Pourquoi s'intéresse-t-on aux jeunes et non à la population entière ? Dans quelle mesure le chômage touche-t-il plus les jeunes que le reste de la population ? Les jeunes sont-ils

réellement plus diplômés ? Pour répondre à ces questions, nous avons dans un premier temps étudié la distribution du chômage et des activités professionnelles au sein de la population Française. En observant les diagrammes circulaires présentés en Figure 1, on constate que les chômeurs représentent 6 % de la population Française en 2013, et que 39 % des chômeurs ont entre 15 et 29 ans. Pourtant, les jeunes entre 15 et 29 ans représentent seulement 19% de la population française.



Afin d'étudier le lien statistique entre "être un chômeur" et "être jeune", nous avons effectué un tableau de contingence entre la variable ACTEU, qui est une variable binaire indiquant si l'individu est actif (0) ou chômeur (1), et la variable AGE3 correspondant à la tranche d'âge de l'individu réparti en trois modalités ("15-29 ans", "30-49 ans" et "+50ans"). Ci-dessous les résultats que nous avons obtenus :

	Age	ACTIF	CHOMEUR	Total ligne
Contenu des cases	15-29ans	31539	7911	39450
		460.645	4244.649	
		0.799	0.201	0.189
		0.167	0.387	
Effectif de la case	30-49ans	98385	8621	107006
		35.640	328.406	
		0.919	0.081	0.513
		0.522	0.422	
Chi-square contribution	+50ans	58422	3908	62330
		85.614	788.899	
		0.937	0.063	0.299
		0.310	0.191	
Rapport effectifs case/ligne	Total col	188346	20440	208786
		0.902	0.098	
Rapport effectifs case/colonne				
Statistique du Chi2				
4653.984				
Degré de liberté				
2				
p-value				
$\leq 0.0009$				

TABLE 1 – Tableau de contingence entre la variable ACTEU et AGE3

La première observation marquante que l'on peut faire sur ce tableau est la sur-représentation des jeunes de 15 à 29 ans parmi les chômeurs. Les jeunes représentent 39% des chômeurs, tandis qu'ils ne représentent seulement que 19% de la population active ou au chômage. De façon similaire, les chômeurs représentent 20% des jeunes de 15 à 29 ans, tandis qu'ils représentent seulement 10% de la population active ou au chômage.

A l'aide de la statistique du Chi2 de chaque cellule, il nous est possible de déterminer quels facteurs influent le plus sur l'écart à la situation d'indépendance entre les deux variables. Plus celle-ci est élevée, plus la cellule en question contribue de manière significative à la corrélation des deux variables. On peut constater que celle-ci est significativement importante et plus élevée (4244.649) pour la cellule regroupant les chômeurs entre 15 et 29 ans : la sur-représentation des jeunes parmi les chômeurs contribue de manière significative à l'écart à la situation d'indépendance des deux variables.

Afin d'estimer si l'association statistique des deux variables est bien significative, il convient de faire un test du Chi2. On obtient une p-value inférieur à 0,001%. Nous pouvons donc rejeter l'hypothèse nulle (ici d'indépendance) en faveur de l'hypothèse de corrélation. Tout mène à penser que les jeunes sont touchés de manière bien plus significative par le chômage que le reste de la population, ce qui souligne l'intérêt d'étudier le chômage dans cette tranche de la population.

Afin de vérifier l'hypothèse selon laquelle les jeunes sont plus diplômés, nous avons effectué un tableau de contingence sur la variable DIP11 correspondant au diplôme le plus élevé que nous avons recodé en 4 modalités (Moins que BAC, BAC, Bac +2, Master/Doctorat), et la variable AGE3. Nous avons mis le tableau en Annexe 3, mais l'analyse des résultats reste la même : les deux variables sont significativement corrélées, avec une sous-représentation des jeunes et une sur-représentation des plus âgés chez les très peu diplômés.

Par la suite, au sein de la population des 15-29 ans, nous allons devoir étudier les sous-échantillons de la population active et passive ensemble et celui des chômeurs seuls séparément. Il serait donc intéressant de comparer les deux sous-échantillons des actifs et des chômeurs. Nous allons observer les répartitions de ceux-ci par rapport à une variable que l'on utilisera dans nos modèles. Il faudrait toutefois qu'elle corresponde à une question qui a été posée aux deux sous-échantillons. La question "quel est votre niveau d'étude?" est satisfaisante, nous sélectionnons donc DIP11 (en 4 modalités correspondantes au niveau d'étude de l'interrogé) pour comparer les sous-échantillons.

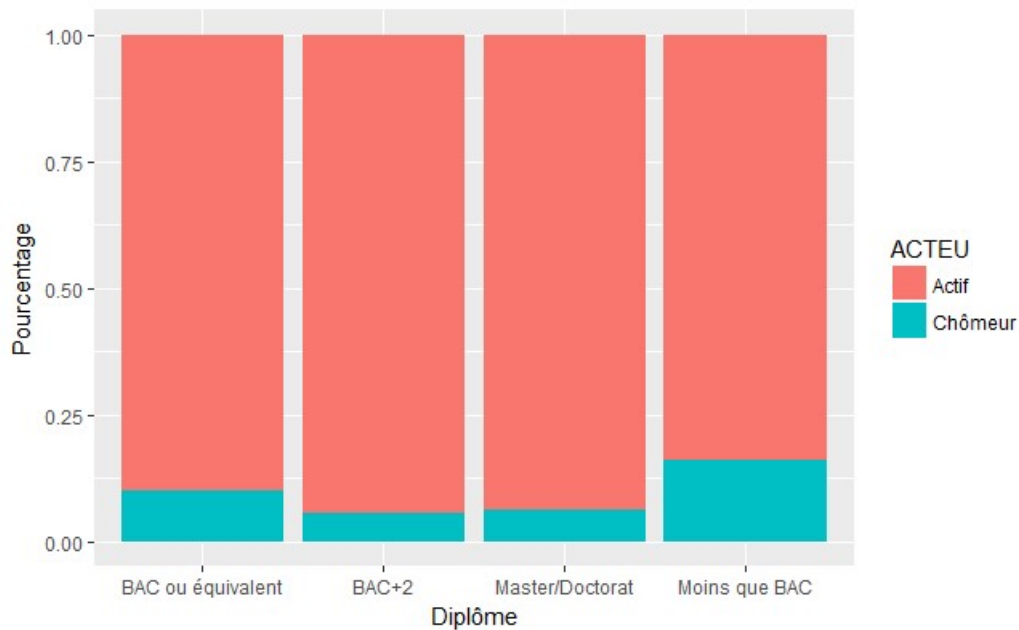


FIGURE 1 – Répartition des actifs et chômeurs pour chaque type de diplôme obtenu

Les différences entre ces deux sous-échantillons sont flagrantes. En plus de représenter 20% des jeunes contre 80% pour les actifs, la répartition des chômeurs est très différente de celle des actifs au sein des diplômés. On remarque que plus le niveau scolaire est bas, plus les chômeurs représentent une grande part des diplômés dans ces tranches. Ainsi, 25% de ceux qui ont moins que le Bac sont des chômeurs et moins de 10% de ceux qui ont un Bac+2 sont chômeurs. Ceci nous donne une idée d'à quel point ces deux sous-échantillons sont différents. De plus, cela nous permet d'émettre une première hypothèse sur le lien entre l'expérience et le chômage chez les jeunes : l'expérience académique influerait de façon négative sur l'emploi chez les jeunes.

### 3 L'influence du diplôme sur le chômage

Dans cette partie, nous avons essayé de quantifier l'influence qu'a le niveau de diplôme sur le niveau d'embauche chez les jeunes. La variable à expliquer ( $Y=ACTEU$ ) étant binaire (1=Actif, 0=chômeur), nous avons choisi un modèle binomial : la régression logistique.

Avant d'aborder le modèle, il convient de vérifier l'intérêt de considérer la variable du diplôme pour expliquer la variable  $ACTEU$  à l'aide d'un test du Chi-2 : On obtient un Chi2 de 1636.571 à 3 degré de liberté ainsi qu'une p-value inférieure à 0.001, ce qui nous permet de rejeter l'hypothèse nulle d'indépendance entre  $ACTEU$  et  $DIP11$  au seuil de 1%.

Nous avons donc choisi une régression logit de l'indicatrice du fait d'être au chômage,  $ACTEU$ , sur  $DIP11$  déterminant le diplôme le plus élevé ainsi que sur la variable  $SEXE$ . La variable  $SEXE$  sert ici principalement comme variable de contrôle. En effet, on pour-

rait penser qu'à diplôme égal, le fait d'être une femme ou un homme peut expliquer en partie le fait d'être chômeur ou actif (Discrimination à l'embauche, niveau de chômage différent dans les secteurs d'activité féminin/masculin, etc.). Nous n'allons cependant pas interpréter les coefficients de la variable SEXE car ils ne nous intéressent pas dans le cadre de notre étude.

Notre modèle est donc :

$$Y^* = X'\beta_0 + \epsilon \text{ où } \epsilon \perp X, \text{ de loi logistique}$$

avec  $Y^*$  la variable latente représentant le niveau "d'employabilité" de l'individu. On observe cependant seulement :

$$ACTEU = \mathbb{1}\{Y^* \geq 0\}$$

On a alors

$$\mathbb{P}(ACTEU = 1|X) = \mathbb{P}(X'\beta_0 + \epsilon \geq 0|X) = \mathbb{P}(-\epsilon \leq X'\beta_0|X) = \frac{1}{1 + \exp(X'\beta_0)}$$

Voici les résultats obtenus :

	Estimate	Std. E	z value	$\mathbb{P}(\geq  z )$	OddRatio	2.5%	97.5%
Intercept	-0.6540	0.027	-24.45	$\leq 2e-16$	0.5199	0.493	0.548
Femme	0.1363	0.026	5.22	$\leq 1.8e-07$	1.1461	1.089	1.206
Bac	-0.7683	0.031	-24.96	$\leq 2e-07$	0.4638	0.437	0.493
Bac +2	-1.4754	0.048	-30.37	$\leq 2e-16$	0.2287	0.208	0.251
Master/Doctorat	-1.5092	0.0458	-32.97	$\leq 2e-16$	0.2211	0.202	0.242

TABLE 2 – Régression logit de la variable SEXE et DIP11 sur la variable ACTEU

Toutes les variables sont significatives. Intéressons nous aux odd-ratios. Les odd-ratios s'obtiennent en prenant l'exponentielle de l'estimé pour chaque variable. En effet, si l'on souhaite calculer l'odd-ratio correspondant à  $X_i = 1$  contre  $X_i = 0$  avec  $X_i$  une variable explicative,  $\beta_i$  le coefficient de  $X_i$ ,  $X_{-i}$  le vecteur contenant toutes les variables explicatives sauf la  $i$ ème et  $\beta_{-i}$  le vecteur contenant les coefficients de  $X_{-i}$ , nous obtenons :

$$\begin{aligned} Oddratio &= \frac{\mathbb{P}(Y = 1|X_i = 1, X_{-i} = x_{-i})}{\mathbb{P}(Y = 0|X_i = 1, X_{-i} = x_{-i})} \times \frac{\mathbb{P}(Y = 0|X_i = 0, X_{-i} = x_{-i})}{\mathbb{P}(Y = 1|X_i = 0, X_{-i} = x_{-i})} \\ &= \frac{[1 + \exp(-\beta_i - x'_{-i}\beta_{-i})]^{-1}}{1 - [1 + \exp(-\beta_i - x'_{-i}\beta_{-i})]^{-1}} \times \frac{[1 + \exp(-x'_{-i}\beta_{-i})]^{-1}}{1 - [1 + \exp(-x'_{-i}\beta_{-i})]^{-1}} \\ &= \frac{\exp(\beta_i + x'_{-i}\beta_{-i})}{\exp(x'_{-i}\beta_{-i})} = \exp(\beta_i) \end{aligned}$$

L'odd-ratio correspondant à la modalité "Master/Doctorat" est compris dans l'intervalle [0.202,0.242] (intervalle de confiance 5%). Un individu ayant un master a donc entre

$\frac{1}{0.242} = 4.13$  et  $\frac{1}{0.202} = 4.95$  fois plus de chance d'être non chômeur qu'un individu chômeur ou actif ayant un diplôme inférieur.

L'hypothèse selon laquelle le diplôme a une influence significative sur l'embauche semble être cohérente. Pour confirmer cette hypothèse, nous avons effectué un F-test pour voir si l'effet global de la variable DIP11 était significative. Nous avons obtenu une p-value inférieure à 0.001, nous pouvons donc rejeter l'hypothèse selon laquelle le diplôme n'a pas d'influence sur l'embauche chez les jeunes.

Les estimés de la modalité "BAC +2" et "Master/Doctorat" étant très proches (respectivement -1.48 et -1.51), nous avons voulu tester si l'effet d'avoir un master était significativement différent de celui d'avoir seulement un BAC +2. Nous avons donc effectué un test-t avec comme hypothèse nulle  $H_0$  "la différence des deux coefficients est égale zéro" et comme hypothèse alternative  $H_1$  "elle est différente de zéro". Nous avons obtenu une p-value de 0.55, ce qui montre que nous ne pouvons pas rejeter l'hypothèse nulle : avoir un master n'apporte pas de manière significative plus de chance d'être embauché.

Ainsi, obtenir des diplômes de qualité croissante permet de diminuer grandement les risques d'être au chômage, ce qui est assez instinctif. L'expérience académique a une véritable influence négative sur le fait d'être au chômage, mais l'expérience d'un individu en général ne se limite pas à son diplôme. Il est aussi intéressant d'étudier l'expérience professionnelle et non plus seulement académique.

## 4 L'effet de l'expérience sur la durée du chômage

L'expérience professionnelle antérieure reflète de notre connaissance "pratique" du monde professionnel, et on peut donc émettre l'hypothèse que c'est un plus pour se faire embaucher durablement. Pour cela, nous allons considérer la variable AAC (1=a déjà eu une expérience professionnelle, 0=pas d'expérience professionnelle antérieure). Toutefois, nous devons nous restreindre à la base des individus au chômage, les actifs ne répondant pas à cette question. La variable d'outcome ACTEU, qui est une indicatrice d'être au chômage ou non n'est donc pas pertinente pour mettre en évidence l'effet de l'expérience professionnelle sur la possibilité d'être au chômage. En effet, elle va être toujours égale à 1 sur notre nouvelle base. Nous allons donc étudier une nouvelle variable caractérisant le chômage, qui va, de plus, nous permettre de faire des observations complémentaires à celles déjà faites préalablement. Cette variable est ANCCCHOM, la durée du chômage en 8 modalités correspondant à des tranches de durée de chômage, dans l'ordre croissant (ce qui nécessitera l'utilisation d'un modèle polytomique). Pour se faire une première intuition de ce que représente la variable, nous l'avons représenté dans la Figure 3 ci-dessous.

Distribution de l'ancienneté de chômage parmi les jeunes chômeurs

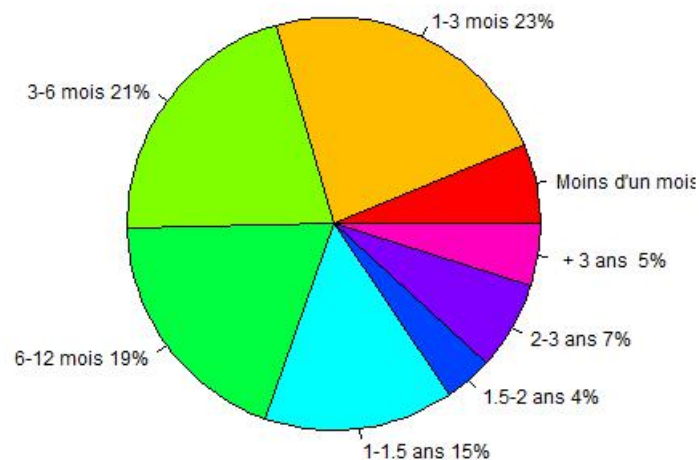


FIGURE 2 – Répartition de l'ancienneté de chômage parmi les chômeurs entre 15 et 29 ans

Un peu moins de la moitié des jeunes chômeurs sont au chômage depuis plus de 6 mois, ce qui est considérable. Nous allons donc, dans cette partie, étudier la durée du chômage et l'influence de l'expérience (professionnelle comme académique) sur celle-ci. Nous régressons par intervalle la variable d'outcome, ANCCCHOM, sur les variables explicatives qui sont DIP11, le diplôme, AAC, l'expérience professionnelle antérieure et la variable de contrôle est encore SEXE. On choisit une régression par intervalle car les seuils de notre variable d'outcome sont connus.

Notons qu'une remarque pourrait être faite sur cette partie : pourquoi ne pas combiner le modèle dichotomique précédent et le polytomique de cette partie en un seul modèle polytomique (être un actif revenant à ANCCCHOM=0)? Cela est dû au fait que nous devons travailler ici sur la base restreinte aux chômeurs, alors que précédemment, nous devons travailler sur les chômeurs aussi bien que les actifs. En effet, nous n'avons pas accès à l'expérience professionnelle antérieure des actifs, AAC ne concernant que les chômeurs. Combiner les deux modèles en un seul est donc impossible.

Avant d'aborder le modèle, nous allons vérifier l'intérêt de considérer la variable AAC, (1=a déjà eu une expérience professionnelle, 0=pas d'expérience professionnelle antérieure). Un test du Chi-2 avec ANCCCHOM montre que cette variable a un certain pouvoir explicatif : le Chi-2 vaut 400 et la p-value, de  $2.07e-85$  nous permet de rejeter l'hypothèse nulle d'indépendance entre AAC et ANCCCHOM. Il est donc pertinent de sélectionner cette variable dans la suite. De même, pour ce qui est de la variable DIP11, on peut raisonnablement supposer son fort pouvoir explicatif de la variable ANCCCHOM, d'après les résultats de la partie précédente. Menons toutefois un test du Chi-2 (la statistique vaut 301.5 et sa p-value vaut  $1.25e-59$ ), qui confirme que DIP11 reste une variable pertinente à étudier, comme précédemment.



Notre modèle est donc :

$$Y^* = X'\beta_0 + \sigma_0\epsilon \text{ où } \epsilon \perp X, \text{ de loi } N(0,1)$$

$Y^*$  est la durée réelle du chômage, mais on observe ANCCCHOM :

$$ANCCCHOM = \sum_{i=1}^8 i. \mathbb{1}\{\alpha_{0i} < Y^* \leq \alpha_{0i+1}\}$$

Les  $\alpha_{0k}$  sont les seuils observés : de "moins d'un mois" à "trois ans ou plus" (8 modalités à ANCCCHOM). Dans notre code, nous avons traduit les durées en mois.

Les résultats de notre régression par intervalle sont les suivants (pour plus de clarté, on a renommé "SEXE" par "femme", indicatrice d'être une femme, et AAC par "travail\_antérieur", indicatrice d'une expérience professionnelle antérieure) :

	Estimate	Std. E	z value	$\mathbb{P}(\geq  z )$	2.5%	97.5%
Femme	-0.579	0.245	-2.36	0.018	-1.06	-0.0985
Travail antérieur	-3.17	0.252	-12.6	0.000	-3.66	-2.68
Bac	-2.84	0.277	-10.3	0.000	-3.38	-2.29
Bac +2	-5.27	0.479	-11.0	0.000	-6.21	-4.33
Master/Doctorat	-6.06	0.453	-13.4	0.000	-6.95	-5.17
Intercept	14.9	0.275	54.0	0.000	14.3	15.4

TABLE 3 – Régression probit ordonnée de ANCCCHOM sur SEXE, DIP11 et AAC

Le Chi-2 du rapport de vraisemblance est de 448.16, et la p-value du Chi-2 à 5 degrés de liberté est quasi nulle. Le modèle est donc globalement significatif. Toutes les variables sont significatives. Interprétons alors les paramètres estimés. Pour ce qui est de l'expérience académique, nos résultats sont en accord avec ceux de la régression logit sur l'indicatrice du chômage, ACTEU. Les paramètres estimés étant négatifs, la qualité du diplôme obtenu a une influence négative sur le temps qu'un individu passe au chômage. Notre modalité de référence pour la variable DIP11 est la première : ne pas avoir le BAC. L'obtention d'un diplôme de qualité croissante provoque une baisse dans la variable expliquée qui est la durée du chômage (ANCCCHOM).

En effet :

- Avoir le BAC par rapport à ne pas l'avoir (toutes autres covariables étant gardées constantes) provoque une baisse de 2.84 mois dans la durée du chômage.
- Pour le BAC+2, cette baisse est encore plus grande : elle est de 5.27 mois par rapport à la modalité de référence.
- Mais la baisse suscitée par le fait d'avoir Master/Doctorat par rapport à ne pas avoir le BAC (toutes autres covariables étant gardées constantes) n'est pas beaucoup plus grande : elle est de 6.06 mois. Avoir un Master/Doctorat baisse la durée de chômage de seulement 1.15 mois par rapport au fait d'avoir un BAC+2.

On avait mis précédemment en évidence que le Master/Doctorat n'apportait pas significativement plus de chances d'être embauché qu'un BAC+2. Les intervalles de confiance

des deux estimateurs correspondant se croisent dans ce modèle, la baisse dans la durée du chômage apportée par un Master/Doctorat par rapport à un Bac+2 n'est pas significative.

Pour ce qui est de l'expérience professionnelle, on remarque que celle-ci a bel et bien une influence négative sur la durée du chômage. Ainsi, un passage de 0 à 1 dans l'indicatrice "travail antérieur", toutes les autres covariables étant maintenues constantes, mène à une baisse de 3.17 mois dans la variable ANCCHOM, la durée du chômage.

Suite à ces observations, on peut aussi affirmer que les effets des variables explicatives considérées sur la durée du chômage sont bel et bien globalement significatifs : pour ce qui est de la variable DIP11, qui a plus de deux modalités, cela est mis en évidence par un test du Chi-2 à trois degrés de liberté. La p-value de ce test est quasi nulle. DIP11 est un prédicateur statistiquement significatif de ANCCHOM, la durée de chômage.

L'expérience professionnelle et académique sont donc bien de véritables atouts pour se faire embaucher le plus vite possible. L'amélioration de cette expérience permet bien de réduire significativement le temps passé au chômage, ces effets se mesurent même en terme de mois.

## 5 Conclusion

Ainsi, nous avons mis en évidence l'importance de l'expérience (professionnelle comme académique) chez les jeunes sur le fait d'être au chômage ou non, mais aussi sur la durée de chômage, en étudiant deux sous-échantillons différents de la population jeune : les actifs-chômeurs ensemble, et les chômeurs seuls.

Nos résultats sont clairs : l'expérience est un facteur déterminant de l'emploi chez les jeunes. L'expérience académique a une influence négative sur le fait d'être au chômage, et les expériences professionnelles comme académiques ont une influence négative sur la durée du chômage si l'on n'est pas employé. Notamment pour l'expérience académique, plus le diplôme est de qualité (plus d'années d'étude), plus l'influence de la modalité en question est négative sur les variables du chômage et de la durée du chômage. On remarque toutefois que l'apport additionnel d'un diplôme de catégorie supérieur diminue passé un certain point. Ainsi, que ce soit pour le fait d'être au chômage ou non, ou pour la durée du chômage, avoir un Master/Doctorat n'a pas un apport statistiquement significatif par rapport à l'obtention d'un BAC+2.

En considérant le coût (temporel et monétaire) des études, il n'est donc pas obligatoirement très rentable de continuer à étudier passé un certain point, si notre objectif est simplement de ne pas être au chômage ou d'y rester le moins longtemps.

Bien entendu, il peut toujours être rentable de maximiser son expérience académique, car nous nous sommes intéressés uniquement au chômage chez les jeunes, alors que l'expérience influe grandement sur le salaire. Il pourrait être intéressant de prolonger cette étude en observant non plus les variables en lien avec le chômage, mais aussi celles en lien avec le salaire, ceci en travaillant sur le sous-échantillon des jeunes actifs.

## Annexe

### 5.1 Annexe 1 : Définition du chômage

Pour être chômeur au sens du BIT, il faut répondre à ses quatre critères :

1. ne pas être en emploi au sens du BIT (tel que défini précédemment)
2. avoir moins de 75 ans
3. être disponible pour travailler dans les quinze jours à venir
4. être dans l'une au moins des trois situations suivantes :
  - avoir effectué des démarches actives de recherche d'emploi dans les quatre semaines s'achevant par la semaine de référence
  - avoir trouvé un emploi qui commence dans moins de trois mois
  - occuper un emploi saisonnier, pour lequel la semaine de référence tombe en morte saison, mais qui (ré)démarre dans moins de trois mois.

### 5.2 Annexe 2 : Liste des variables

**ACTEU** : Statut d'activité (Chômeur, actif ou inactif)

**AGE3** : Age au dernier jour de la semaine de référence (15-30, 30-49, 50+)

**ANCCHOM** : Ancienneté du chômage en 8 postes

**DIP11** : Diplôme le plus élevé obtenu (2 chiffres, 11 postes)

**EXTRIAN14** : Coefficient annuel de pondération des individus

**MRBBIS** : Mode de recherche active d'emploi durant les quatre semaines s'achevant par la semaine de référence : a déjà fait des missions avec une agence d'intérim

**SEXE** : Genre de l'individu

**SITANT** : Situation professionnelle antérieure à l'emploi occupé

**AAC** : Exercice d'une activité professionnelle régulière antérieure, pour les inactifs, chômeurs et personnes ayant une activité temporaire ou d'appoint autre qu'un emploi informel

### 5.3 Annexe 3 : Tableau de contingence Age/diplôme

Age	Moins que BAC	BAC	BAC+2	master/doctorat	Row Total
15-29ans	7025	19421	5810	7080	39336
	205.888	119.169	0.882	0.914	
	0.179	0.494	0.148	0.180	0.189
	0.159	0.204	0.186	0.187	
	0.034	0.093	0.028	0.034	
30-49ans	18174	47858	18624	22167	106823
	879.176	16.981	439.795	380.587	
	0.170	0.448	0.174	0.208	0.513
	0.412	0.503	0.598	0.584	
	0.087	0.230	0.089	0.106	
+50ans	18950	27842	6722	8683	62197
	2527.024	10.765	714.859	615.362	
	0.305	0.448	0.108	0.140	0.299
	0.429	0.293	0.216	0.229	
	0.091	0.134	0.032	0.042	
Column Total	44149	95121	31156	37930	208356
	0.212	0.457	0.150	0.182	

Chisq	D.F	p.value
5661.13	6	0.000

### 5.4 Annexe 4 : Code Stata

```

1 use "\\paradis\eleves\FTRAINAR\Bureau\df_chom3.dta", clear
2
3 gen borne_sup = 1*(ANCCHOM=="1")+3*(ANCCHOM=="2")+6*(ANCCHOM=="3")+12*(
  ANCCHOM=="4")+18*(ANCCHOM=="5")+24*(ANCCHOM=="6")+36*(ANCCHOM=="7")
  if ANCCHOM != "8" & ANCCHOM != "VIDE"
4 gen borne_inf = 1*(ANCCHOM=="2")+3*(ANCCHOM=="3")+6*(ANCCHOM=="4")+12*(
  ANCCHOM=="5")+18*(ANCCHOM=="6")+24*(ANCCHOM=="7")+36*(ANCCHOM=="8")
  if ANCCHOM != "1" & ANCCHOM != "VIDE"
5
6
7 gen femme = (SEXE=="1")
8 gen travail_anterieur = (AAC=="1")
9 destring(DIP11), replace
10
11
12 intreg borne_inf borne_sup femme travail_anterieur i.DIP11
13
14 margins DIP11

```

## 5.5 Annexe 4 : Code R

```

1 library(foreign)
2 library(data.table)
3 library(reshape2)
4 library(questionr)
5 library(mfx)
6 library(gmodels)
7 library(aod)
8 library(ggplot2)
9 library(weights)
10 library(MASS)
11 library(survival)
12 #creation des dataframe
13 df <- read.dbf("C:\\Users\\maria\\Documents\\ENSAE\\Projet\\Econometrie
    \\eec13.dbf")
14 varlist <- read.dbf("C:\\Users\\maria\\Documents\\ENSAE\\Projet\\
    Econometrie\\varlist.dbf")
15 varmod <- read.dbf("C:\\Users\\maria\\Documents\\ENSAE\\Projet\\
    Econometrie\\varmod.dbf")
16 varlist["MODALITE"]<-NA
17 row.names(varlist)<-varlist[["VARIABLE"]]
18 for (i in as.character(varlist[["VARIABLE"]][,])){
19   varlist[i,"MODALITE"]<-toString(as.character(varmod[varmod[["VARIABLE"]]=i,][["MODALITE"]]))
20 }
21 var1 = c("ACTEU","AGE3","ANCCHOM","CONTACT","CSA","CSER","DEM","DIP11",
    "ENFRED","EXTRIAN14","HREC","MRC","MRD","NAFANTG004","NFRRED","NREC",
    "NRECA","NRECB","OFFICC","PERCREV","RAISPAS","SEXE","SITANT","
    TYPMEN7","MRBBIS","SOUSEMPL","SITANT","AAC","MRBBIS","COURED")
22 var2 = c("COURED","DISPOC","STCHE")
23 row.names(df)=1:nrow(df)
24 df=df[,var1]
25 varlist=varlist[var1,]
26 is.na.data.frame <- function(x)
27 do.call(cbind, lapply(x, is.na))
28 df[sapply(df, is.factor)] <- lapply(df[sapply(df, is.factor)], as.
    character)
29 df[is.na(df)] <- "VIDE"
30 jpeg('Activite_population.jpg')
31 slices <- c(colSums(df[df["ACTEU"]==1,][["EXTRIAN14"]]),colSums(df[df["
    ACTEU"]==2,][["EXTRIAN14"]]),
32           colSums(df[df["ACTEU"]==3,][["EXTRIAN14"]]))
33 lbls <- c("Actif Occupe", "Chomeur", "Inactif")
34 pct <- round(slices/sum(slices)*100)
35 lbls <- paste(lbls, pct)
36 lbls <- paste(lbls, "%", sep="")
37 pie(slices, labels = lbls, main="Repartition de l'activite au sein de
    la population",col=rainbow(length(lbls)))
38 dev.off()
39 pie(slices, labels = lbls, main="Repartition de l'activite au sein de
    la population",col=rainbow(length(lbls)))
40 jpeg('Anciennete_de_chomage.jpg')
41 df_chom=df[df["ACTEU"]==2,]
42 list<-c()

```

```

43 for (i in 1:8) list<-c(list,colSums(df_chom[df_chom["ANCCHOM"]==i,][ "
    EXTRIAN14" ]))
44 \endcsname
45 slices <- list
46
47 lbls <- c("Moins d'un mois", "1-3 mois", "3-6 mois", "6-12 mois", "1-1.5
    ans", "1.5-2 ans", "2-3 ans", " + 3 ans ")
48 pct <- round(slices/sum(slices)*100)
49 lbls <- paste(lbls, pct)
50 lbls <- paste(lbls, "%", sep="")
51 pie(slices, labels = lbls, main="Distribution de l'anciennete de
    chomage parmi les chomeurs",
52     col=rainbow(length(lbls)))
53 dev.off()
54 pie(slices, labels = lbls, main="Distribution de l'anciennete de
    chomage parmi les chomeurs",
55     col=rainbow(length(lbls)))
56 jpeg('Age-chomeur.jpg')
57 list<-c()
58
59 slices <- c(colSums(df_chom[df_chom["AGE3"]==15,][ "EXTRIAN14" ]),colSums
    (df_chom[df_chom["AGE3"]==30,][ "EXTRIAN14" ]),
60     colSums(df_chom[df_chom["AGE3"]==50,][ "EXTRIAN14" ]))
61 lbls <- c("15-29 ans", "30-49 ans", " + 50 ans")
62 pct <- round(slices/sum(slices)*100)
63 lbls <- paste(lbls, pct)
64 lbls <- paste(lbls, "%", sep="")
65 pie(slices, labels = lbls, main="Repartition de l'age parmi les
    chomeurs",col=rainbow(length(lbls)))
66 dev.off()
67 pie(slices, labels = lbls, main="Repartition de l'age parmi les
    chomeurs",col=rainbow(length(lbls)))
68
69 df2=df[(df["ACTEU"]==1)|(df["ACTEU"]==2),]
70
71
72 table=CrossTable(factor(df2$AGE3, levels=c(15,30,50), labels=c("15-29ans"
    , "30-49ans", "+50ans")),
73     factor(df2$ACTEU, levels=c(1,2), labels=c("ACTIF", "
    CHOMEUR")), weight=df2$EXTRIAN14,
74     dnn=c("Age", "Activite"), plot=TRUE)
75
76
77 wtd.chi.sq(df2$AGE3, df2$ACTEU, weight=df2$EXTRIAN14)
78 df2$ACTEU[df2$ACTEU==1]<-0
79 df2$ACTEU[df2$ACTEU==2]<-1
80 df2$ACTEU<-as.numeric(df2$ACTEU)
81
82 df2$DIP11[df2$DIP11 %in% c(10,11)]<-4#"master/doctorat"
83 df2$DIP11[df2$DIP11 %in% c(33,31,30)]<-3#"BAC+2"
84 df2$DIP11[df2$DIP11 %in% c(41,42,50)]<-2#"BAC ou equivalent"
85 df2$DIP11[df2$DIP11 %in% c(60,70,71)]<- 1#"Moins que BAC"
86 df2<-df2[df2$DIP11 != "VIDE",]
87 df2$SEXE[df2$SEXE==1]<-0
88 df2$SEXE[df2$SEXE==2]<-1
89

```

```

90
91 table=CrossTable( factor( df2$AGE3, levels=c(15,30,50), labels=c("15-29ans"
92 , "30-49ans", "+50ans")),
93                     factor( df2$DIP11, levels=c(1,2,3,4), labels=c("Moins que
94 BAC", "BAC ou equivalent", "BAC+2", "master/doctorat
95 ")), weight=df2$EXTRIAN14, dnn=c("Age", "DIPLOME"),
96 plot=TRUE)
97
98 wtd.chi.sq( df2$AGE3, df2$DIP11, weight=df2$EXTRIAN14)
99 df3<- df2[ df2$AGE3==15,]
100 mylogit=glm(ACTEU ~SEXE+DIP11, family=binomial, data=df3)
101 summary(mylogit)
102 #commentaire: Avoir un diplome BAC+2 comparait a juste BAC change le
103 log odds d'etre au chomage de -0.7, critere d'AKAKE: 37812
104 #intervalle de confiance+odd ratio
105 exp(cbind(OR = coef(mylogit), confint(mylogit)))
106 wald.test(b = coef(mylogit), Sigma = vcov(mylogit), Terms = 3:5)
107 l <- cbind(0, 0, 0, 1,-1)
108 wald.test(b = coef(mylogit), Sigma = vcov(mylogit), L = 1)
109 df_chom2=df3[ df3["ACTEU"]==1,]
110 #recodage de AAC ie acti pro anterieure reguliere : 0 =non, 1 = oui
111 df_chom2$AAC[df_chom2$AAC==2]<-0
112 df_chom2$AAC[df_chom2$AAC==1]<-1
113 head(df_chom2)
114 #codons ANCCCHOM2 dans df2_chom, la base des chomeurs et dans df3 pour
115 avoir trois modalites de durees de chomage
116 df_chom2$ANCCCHOM2[df_chom2$ANCCCHOM<=3]<-1
117 df_chom2$ANCCCHOM2[df_chom2$ANCCCHOM<=5 & df_chom2$ANCCCHOM<3]<-2
118 df_chom2$ANCCCHOM2[df_chom2$ANCCCHOM>5]<-3
119 df_chom2$ANCCCHOM2[df_chom2$ANCCCHOM=="VIDE"]<-"VIDE"
120
121 df3$ANCCCHOM2[df3$ANCCCHOM<=3]<-1
122 df3$ANCCCHOM2[df3$ANCCCHOM<=5 & df3$ANCCCHOM<3]<-2
123 df3$ANCCCHOM2[df3$ANCCCHOM>5]<-3
124 df3$ANCCCHOM2[df3$ANCCCHOM=="VIDE"]<-"VIDE"
125 df_chom2[is.na(df_chom2)] <- "VIDE"
126 df3[is.na(df3)] <- "VIDE"
127
128 wtd.chi.sq(df_chom2$AAC, df_chom2$ANCCCHOM2, weight=df_chom2$EXTRIAN14)
129 wtd.chi.sq(df_chom2$DIP11, df_chom2$ANCCCHOM2, weight=df_chom2$EXTRIAN14)
130 df_chom3<-df_chom2[ df_chom2["AAC"]!="VIDE",]
131
132 #exporter la base pour traiter la regression sous stata
133
134 setwd("C:\\Users\\ftrai\\Desktop\\ENSAE\\Projet")
135 getwd()
136 write.dta(df_chom3, file = 'df_chom3.dta')
137 # #Analyse descriptive pour la difference actif/chomeurs
138
139 library(scales)
140 base<-df2[,c("DIP11", "EXTRIAN14", "ACTEU")]
141 for (i in 1:4){base$DIP11[base$DIP11==i]<-lbls[i]}
142
143 base$ACTEU[base$ACTEU==1]<-"Chomeur"
144 base$ACTEU[base$ACTEU==0]<-"Actif"

```

```
140 |  
141 | ggplot(base , aes(x=DIP11 , y=EXTRIAN14, fill=ACTEU)) + geom_bar(position  
    | = "fill" , stat="Identity") + scale_colour_identity() + ylab("Pourcentage"  
    | ) + xlab("Diplome")
```