

## Le programme d'investissement de l'État français (IDEX) a-t-il une influence sur le domaine de recherche des bénéficiaires de la prime d'excellence scientifique ?

GNAGO YANNICK    yannick.gnago@univ-reims.fr  
AA ICHATOU SOW    aa-ichatou-m.sow@univ-reims.fr

14 mai 2021

Appréciations et note du correcteur :

## Résumé

On essayera de mettre en évidence l'influence du programme d'investissement de l'État français dans l'enseignement supérieur et la recherche (IDEX) sur le domaine de recherche des bénéficiaires de la prime d'excellence scientifique. Nous avons donc une base de données relative aux bénéficiaires de la prime d'excellence scientifique de 1993 à 2012 sur-le-site. On a d'abord fait le pré-traitement de la base, puis nous avons étudié quelques variables et ensuite on a fait le croisement des variables afin d'avoir une appréhension sur leurs relations.

## Table des matières

<b>1 L'étude univarié</b>	<b>2</b>
1.1 Le programme d'investissement. <b>idex</b> . . . . .	3
1.2 Le sexe des bénéficiaires. <b>sexe</b> . . . . .	3
1.3 Le domaine de recherche des bénéficiaires. <b>seteur_disciplinaire</b> . . . . .	3
1.4 La profession des bénéficiaires. <b>groupe_de_corps</b> . . . . .	4
1.5 La région des bénéficiaires. <b>region</b> . . . . .	4
1.6 Le lien entre le pôle de de recherche et l'établissement. <b>distance_univ</b> . . . . .	5
<b>2 L'étude bivarié</b>	<b>6</b>
2.1 La répartition entre le sexe et la professions des bénéficiaires. . . . .	6
2.2 La répartition entre le programme d'investissement et le nombre de bénéficiaires. . . . .	7
2.3 La répartition entre le domaine de recherche et le nombre de bénéficiaires. . . . .	7
2.4 La répartition entre le programme d'investissement et le domaine de recherche. . . . .	8
<b>3 Analyse de la Problématique.</b>	<b>10</b>
<b>A Annexes</b>	<b>10</b>
A.1 Le Pré-traitement . . . . .	10
A.2 Test d'indépendance entre le sexe et la professions des bénéficiaires. . . . .	11
A.3 Test d'indépendance entre le programme d'investissement et le nombre de bénéficiaires. . . . .	11
A.4 Test d'indépendance entre le le domaine de recherche et le programme d'investissement. . . . .	12
A.5 Test d'indépendance entre le programme d'investissement et la position du laboratoire de recherche à l'établissement. . . . .	13
A.6 Test d'indépendance entre le sexe et le nombre de bénéficiaires. . . . .	14
<b>B Bibliographie</b>	<b>16</b>

## 1 L'étude univarié

Dans cette partie, nous essayerons au mieux d'expliquer les différentes variables et nous étudierons quelques-unes. Après le pré-traitement, notre jeu de donnée est constitué de 11 variables, dont voici ci-dessous quelques définitions :

- **annee**: l'année d'obtention de la prime (de 1993 à 2012).
- **sexe**: le sexe des bénéficiaires ( **femmes** et **hommes**).

- `seteur_disciplinaire`: le secteur de section CNU(conseil national des universités) du bénéficiaire de la prime i.e. son domaine de recherche. Il y a dans l'ensemble le **Droit et sciences économiques**, les **Lettres et sciences humaines**, la **Santé**, les **Sciences** et les **STAPS**.
- `etablissement`: l'école des bénéficiaires.
- `groupe_de_corps`: la profession du bénéficiaire de la prime. Il y a : des **Professeurs et assimilés**, des **Maîtres de conférences et assimilés**, des **particuliers**(case vide dans la dataframe).
- `region` : la région du laboratoire de recherche des bénéficiaires de la prime.
- `beneficiaires` : Le nombre de bénéficiaires selon l' **etablissement**, le **sexe**, l' **année**, ....
- `distance_univ`: Le lien entre le pôle de de recherche et un etablissement supérieur. Il y a **PRES**<sup>1</sup> (pour les bénéficiaires dont le pôle de recherche est rattaché à l'établissement) et **HORS PRES**(pour les bénéficiaires dont le pôle de recherche n'est pas rattaché en dehors de l'établissement).
- `ideX`<sup>2</sup>: il y a **HORS IDEX** pour les bénéficiaires sans le programme IDEX et **IDEX** pour les bénéficiaires avec le programme IDEX.

### 1.1 Le programme d'investissement. idex

Il représente ici l'effectif des bénéficiaires suivant un le programme d'investissement IDEX ou pas. Parmi les établissements des bénéficiaires primés, on recense 75.4% **HORS IDEX** et 24.6% **IDEX**.

La répartition du programme d'investissement est illustrée par la **figure 1**.

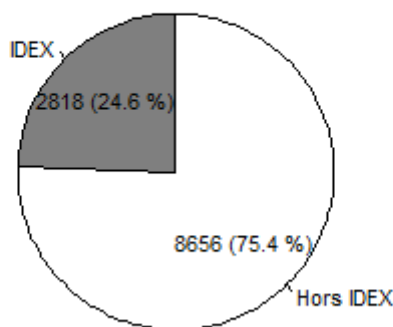


FIGURE 1 – Répartition des établissements selon le statut IDEX/HORS IDEX.

### 1.2 Le sexe des bénéficiaires. sexe

Il représente l'effectif du sexe suivant les bénéficiaires de la prime d'excellence.

Parmi les bénéficiaires, on recense 62.9% d'hommes et 37.1% de femmes durant la période de 1993 à 2012(Voir la figure 2).

### 1.3 Le domaine de recherche des bénéficiaires. seteur\_disciplinaire

Il y a près de 50 % des bénéficiaires qui sont issus du domaine de la science, ensuite 21.9 % sont les lettres et sciences humaines (Voir la figure 3).

Il semble alors que les chercheurs en sciences sont les plus primés, cela est-il dû au programme d'IDEX? (Voir dans l'analyse de la problématique)

---

1. PRES sont les pôles de recherche et d'enseignement supérieur.

2. L'IDEX est un programme d'investissement de l'État français dans l'enseignement supérieur et de recherche.

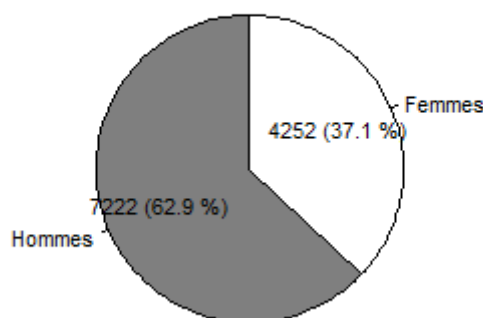


FIGURE 2 – Répartition du sexe chez les bénéficiaires du prix d'excellence scientifique de 1993 à 2012.

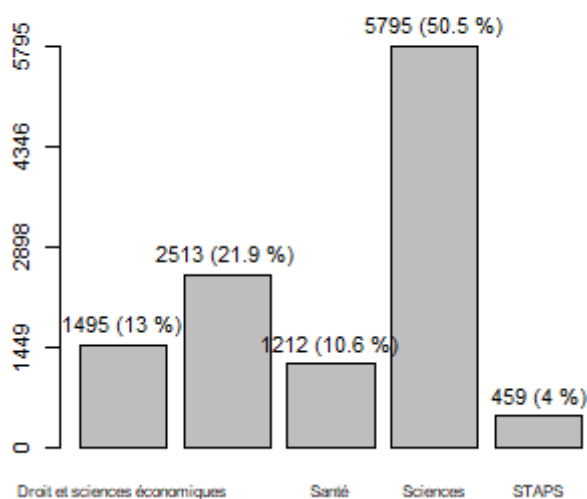


FIGURE 3 – Diagramme en barre des domaines des bénéficiaires au prix d'excellence scientifique de 1993 à 2012.

#### 1.4 La profession des bénéficiaires. groupe\_de\_corps

Nous constatons qu'il y a 63.8 % des bénéficiaires qui sont des Professeurs, 36.0 % Maîtres de conférences et 0.2 % particuliers. Dans l'ensemble, la majorité des bénéficiaires est dans le corps de l'enseignement supérieur, ce qui est d'autant plus logique. (**Voir la figure 4**)

#### 1.5 La région des bénéficiaires. region

La région des bénéficiaires la plus représentée est l'Île de France, puis celle de Rhône-Alpes (**Voir figure 5**). Cela s'explique par le fait qu'elles sont des régions de France avec le plus d'étudiants chercheurs d'après le site et qu'elles représentent les deux plus grandes régions contenant le plus de **PRES**<sup>3</sup> d'après le site.

---

3. PRES sont les pôles de recherche et d'enseignement supérieur.

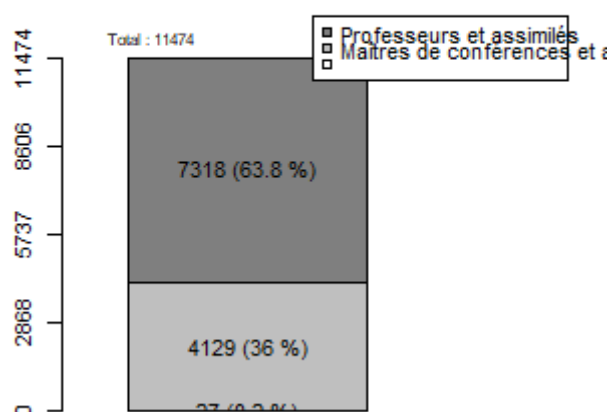


FIGURE 4 – Diagramme en Bande de la profession des bénéficiaires primés au prix d'excellence scientifique.

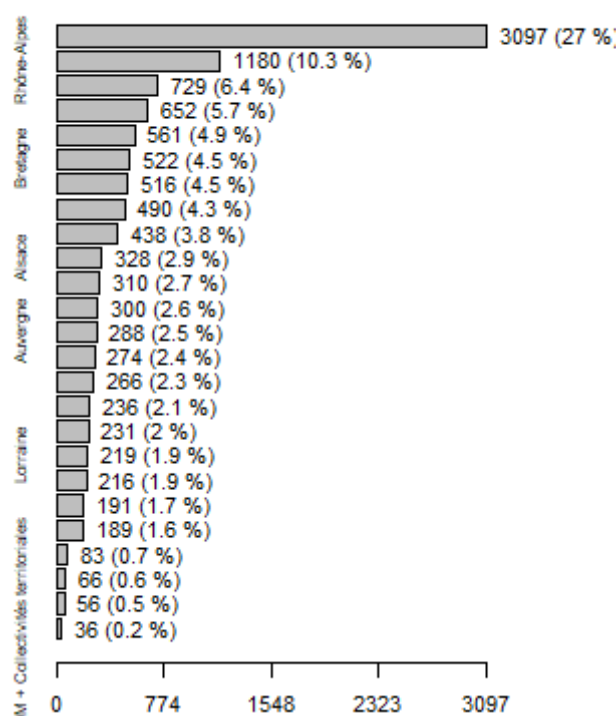


FIGURE 5 – Diagramme en barre représentant les régions des bénéficiaires.

## 1.6 Le lien entre le pôle de de recherche et l'établissement. distance\_univ

On observe qu'il y a environ 85 % des laboratoires de recherche qui sont rattachés aux établissements (Voir la figure 6), cela s'explique peut-être par le fait qu'il y a beaucoup de centres de recherche qui sont en partenariat avec les établissements publics comme la CNRS, l'EPST, l'INRA, l'INRIA, l'INSERM. ....

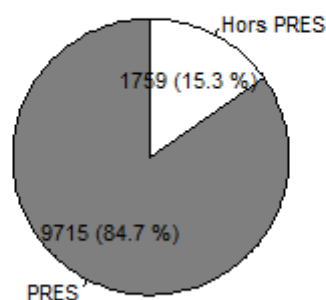


FIGURE 6 – Répartition des bénéficiaires primés en fonction du rattachement du laboratoire de recherche et l'établissement.

## 2 L'étude bivarié

Dans cette seconde partie, nous allons étudier le croisement de certaines variables que nous avons choisi. Parmi ces variables, nous avons les croisements suivantes :

- `eff_sexe_group` : croisement entre les variables Sexe et la professions des bénéficiaires.
- `moy_idx_beneficiaire` : croisement entre les variables du programme d'investissement et le nombre de bénéficiaires.
- `moy_sect_beneficiaire` : croisement entre les variables du domaine de recherche et le nombre de bénéficiaires.
- `eff_idx_secteur` : croisement entre les variables du programme d'investissement et le domaine de recherche.

### 2.1 La répartition entre le sexe et la professions des bénéficiaires.

#### 2.1.1 Profil ligne

L'étude du profil ligne de ces deux variables nous montre que parmi les **femmes** il n'y a pas de **particulier** alors que tel n'est pas le cas pour les **hommes** qui représentent 0.3 %. Cependant, parmi les **maîtres de conférences**, les **femmes** constituent 39.6 %, et 33.9 % sont des **hommes**. Par ailleurs, les **femmes** représentent 60.4 % des **Professeurs** et les **hommes** 65.8 %. (**voir figure 7**)

#### 2.1.2 Profil colonne

Ici, on voit que parmi les **particuliers**, 96.3 % sont des **hommes** alors que les **femmes** ne représentent que 3.7 %. Concernant les **maîtres de conférences**, nous avons 40.8 % de **femmes** et 59.2 % d' **hommes** ; ce qui ne sont pas le cas pour les **Professeurs**, là où 64.9 % sont des **hommes** et les **femmes** ne sont que 35.1 %.(**voir figure 8**)

D'après ces graphes, on voit qu'il y a une différence de résultats entre les partitions dans chaque profil.

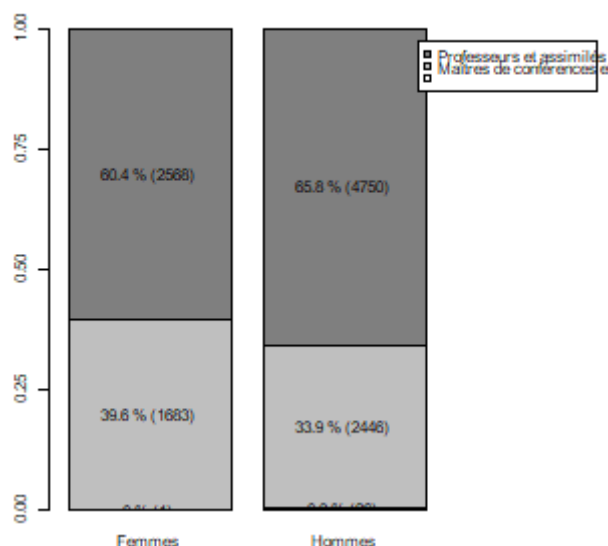


FIGURE 7 – Profil ligne des variables du sexe et la profession représentées sous forme de diagramme en barres

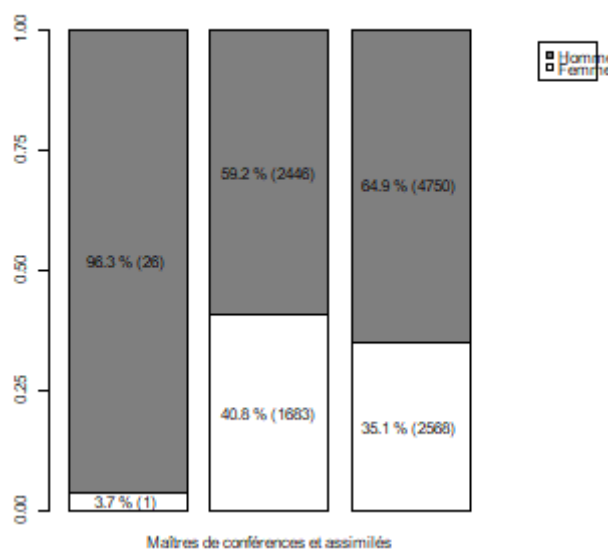


FIGURE 8 – Profil colonne des variables du sexe et la profession représentées sous forme de diagramme en barres

## 2.2 La répartition entre le programme d'investissement et le nombre de bénéficiaires.

Au total 34 231 des bénéficiaires qui ne sont pas suivi par le programme IDEX et 16 667 qui le sont. Il y'a donc plus de bénéficiaires **HORS IDEX** que de bénéficiaires suivant le programme d'investissement **IDEX** . On constate aussi la présence de valeurs aberrantes, une dissymétrie au niveau des boîtes, et un éloignement des moyennes aux médianes. (voir figure 9)

## 2.3 La répartition entre le domaine de recherche et le nombre de bénéficiaires.

On constate qu'il y a plus de bénéficiaires en sciences que dans les autres domaines.(Voir figure 10) On voit qu'en moyenne le nombre de bénéficiaires du secteur **droit et sciences économiques** est de 2.7843101, 4.2668332 pour le secteur **Lettres et sciences humaines**, 2.2333568 pour le secteur **Santé**, 9.5970784

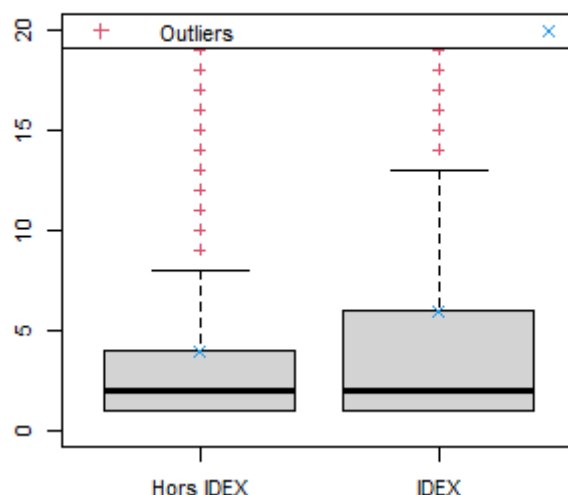


FIGURE 9 – Représentation de la boîte à moustaches juxtaposées du couple iDex/le nombre de bénéficiaires

pour le secteur **Sciences**, et enfin 0.5816679 pour secteurs **STAPS**. Le domaine le moins représenté à la prime d'excellence scientifique est alors les **STAPS**.

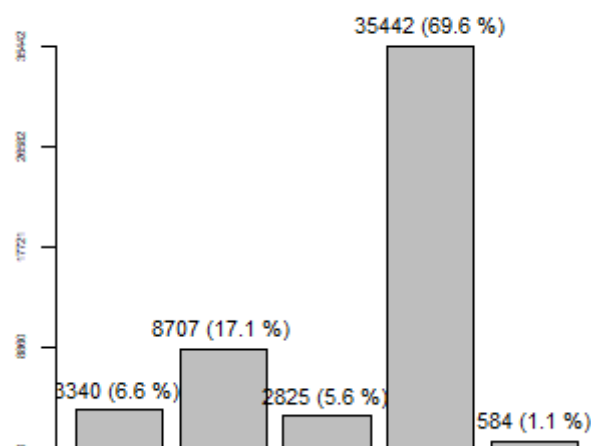


FIGURE 10 – Distributions des bénéficiaires des différents domaines de recherche représentées sous forme de diagramme en barre.

On constate aussi la présence de valeurs aberrantes, une dissymétrie des boîtes, et l'éloignement des moyennes aux médianes. (**voir figure 11**)

## 2.4 La répartition entre le programme d'investissement et le domaine de recherche.

### 2.4.1 Profil ligne

L'étude du profil ligne de ce croisement montre que parmi les **Hors IDEX** il y a 12.6 % en **droit et sciences économiques**, 21.6 % en **lettres et sciences humaines**, 10.0 % en **santé**, 51.7 % en **Sciences**, et 4.1 % en **STAPS**. Et parmi ceux dans **IDEX** il y a 14.5 % en **droit et sciences économiques**, 22.9



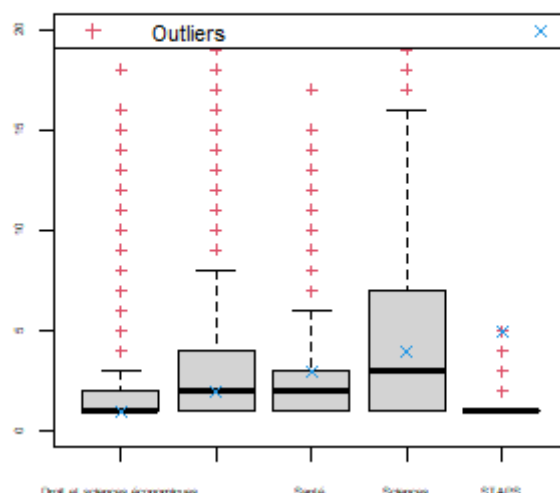


FIGURE 11 – Réprésentation des boîtes à moustaches juxtaposées du couple du domaine/le nombre de bénéficiaires

% en lettres et sciences humaines, 12.2 % en santé, 46.8 % en Sciences, et 3.6 % en STAPS. (voir figure 12)

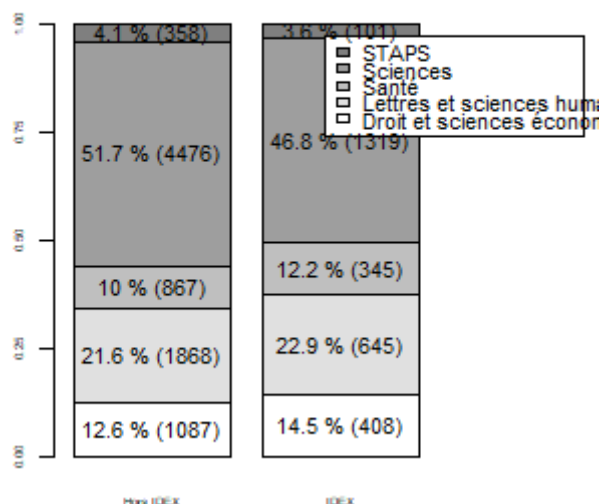


FIGURE 12 – Profil ligne représentée des variables sexes et secteur sous forme de diagramme en barres

## 2.4.2 Profil colonne

Ici, on voit que pour le secteur **Droit et sciences économiques** 72.7 % sont **Hors IDEX** et 27.3 % sont **IDEX**, dans le secteur **Lettres et sciences humaines** il y a 74.3 % sont **Hors IDEX** et 25.7 % **IDEX** , dans le secteur **santé** 71.5% sont **Hors IDEX** et 28.5 % sont **IDEX** , dans le secteur **Sciences** 77.2 % sont **Hors IDEX** et 22.8 % sont **IDEX**, dans le secteur **STAPS** 78 % sont **Hors IDEX** et 22 % sont **IDEX** . (voir figure 13)

D'après ces graphes, on voit qu'il n'y a pas vraiment de différence de résultats entre les partitions des profils.

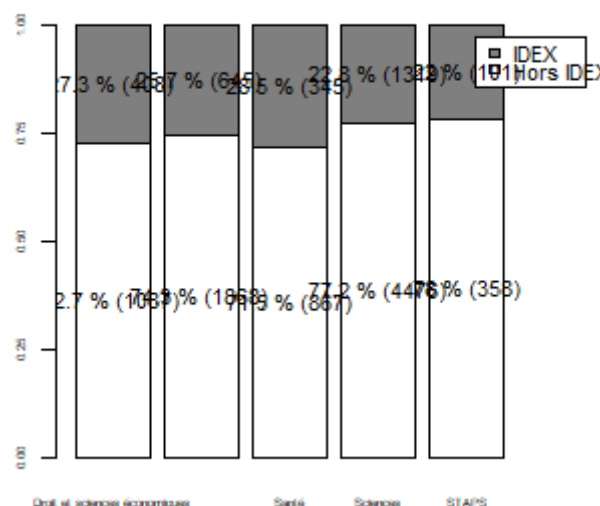


FIGURE 13 – Profil colonne représentée des variables sexes et secteur sous forme de diagramme en barres

### 3 Analyse de la Problématique.

Après l'étude bivarié, nous avons pu observer qu'il semble avoir une relation de dépendance entre le programme d'investissement (IDEX) et le domaine de recherche des bénéficiaires de la prime.

En effet, on peut s'apercevoir, qu'il semble avoir une forte influence entre le programme IDEX et les domaines de santé, de Droit et sciences économiques; ce qui n'est le cas pour le domaine de la science (Voir figure 15). Cela pourrait peut-être s'expliquer par le fait qu'il y a plus de bénéficiaires qui sont **HORS IDEX** et près de la moitié des bénéficiaires qui sont en **sciences** alors pour répondre à la question posée dans 1.3, nous pourrions dire que le programme IDEX n'est pas la véritable raison du nombre élevé de bénéficiaires en science, car la relation du couple (IDEX, sciences) est sous représentée par rapport à la situation théorique de dépendance (Voir figure 15).

Enfin, nous avons pu mettre en évidence que l'influence qu'a le programme d'investissement IDEX sur quelques domaines des bénéficiaires de la prime d'excellence scientifique grâce aux données des bénéficiaires de la prime d'excellence scientifique de 1993 à 2012, mais nous ne pouvons pas affirmer avec certitude cette influence sachant que les résultats trouvés ne reflètent pas l'évolution de nombres de bénéficiaires par rapport au programme IDEX au fil des années.

Cependant, il serait intéressant de traiter ces données par année afin d'avoir une étude plus pertinente sur l'influence d'un programme d'investissement dans la recherche sur le nombre de bénéficiaires de la prime d'excellence scientifique.

## A Annexes

### A.1 Le Pré-traitement

Il a consisté à rendre la base plus compréhensible pour faciliter notre analyse. Nous avons donc supprimé les colonnes référençant certains codes qui ne nous serviront pas (code\_secteur, code\_region, code\_pres, ...), ensuite nous avons ajouté les variables **ideX** et **distance\_univ** en se basant déjà sur la colonne "idex" et "pres" pour les rendre plus exploitables. Ce pré-traitement nous a permis d'avoir une dataframe plus compréhensible que la base de données initiale. La difficulté rencontrée dans cette partie a été de pouvoir modifier les cases vides dans les variables **groupe\_de\_corps** par **particuliers**; car lorsqu'on essaye, toutes les lignes des cases vides sont remplacées par **particuliers** ce qui n'était pas cohérent alors finalement, on

a laissé ces cases vides toutes en supposant qu'ils n'avaient pas de métiers ces bénéficiaires d'où le nom **particuliers** pour les cases vides de **groupe\_de\_corps** dans nos analyses.

## A.2 Test d'indépendance entre le sexe et la professions des bénéficiaires.

Etant donné que ces deux variables sont qualitatives, appliquons la procédure `chisq.test` du logiciel R au tableau des effectifs croisés du couple (sexe, `groupe_de_corps`).

```
## chisq.test(eff_sexe_group) \
Pearson's Chi-squared test \
data: eff_sexe_group \
X-squared = 49.277, df = 2, p-value = 0.0000000001994. ##
```

Au niveau de significativité de 5 %, on conclut qu'il existe un lien de dépendance significatif entre le sexe et les professions des bénéficiaires. L'examen des résidus standardisés (**voir la figure 14**) permet de clarifier que les principales contributions à l'inertie sont dues aux couples (femmes, particuliers), (hommes, Maîtres de conférences et assimilés), (femmes, Professeurs et assimilés) respectivement sous-représentés et (hommes, particuliers), (femmes, Maîtres de conférences et assimilés), (hommes, Professeurs et assimilés) respectivement sur-représentés par rapport à la situation théorique de dépendance.

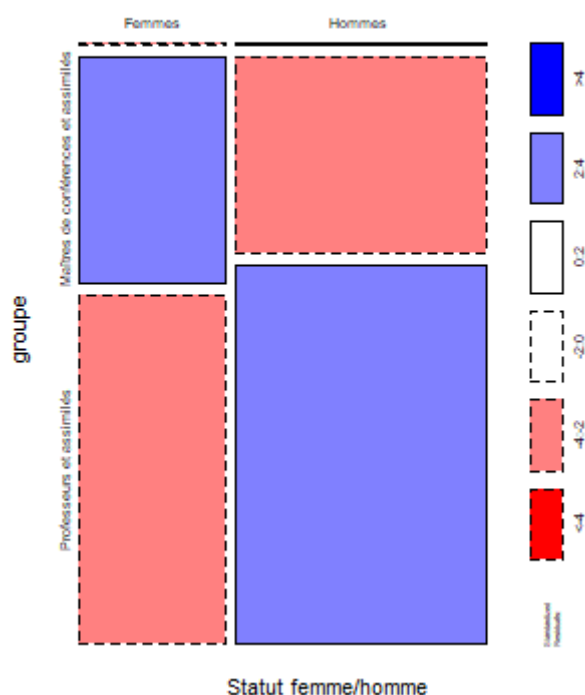


FIGURE 14 – Visualisation des résidus associés entre le sexe et la profession des bénéficiaires.

## A.3 Test d'indépendance entre le programme d'investissement et le nombre de bénéficiaires.

En croisant ces deux variables, nous constatons qu'il y a plus de bénéficiaires **HORS IDEX** que **IDEX**, puisque la variable **beneficiaire** est quantitative et **ideX** qualitative à deux modalités, nous avons donc fait le test de student.

```
## F test to compare two variances \
```

```
data:/quad beneficiaires by ideX\
F = 0.42592, num df = 8655, denom df = 2817, p-value < 0.00000000000000022 \
alternative hypothesis: true ratio of variances is not equal to 1 \
95 percent confidence interval: \
0.4008785 0.4521110 \
sample estimates: \
ratio of variances \

0.4259233 $$
```

On constate une différence notable de variances au sein des sous-populations et la  $p\text{-value} < \text{seuil}$ , on conclut qu'il n'y a pas une différence significative de variance entre les deux populations.

```
$$ Welch Two Sample t-test \
data:/quad beneficiaires by ideX \
t = -9.9461, df = 3629.7, p-value = 1 \
alternative hypothesis: true \
difference in means is greater than 0 \
95 percent confidence interval: \
-2.28408 Inf \
sample estimates: \
mean in group Hors IDEX mean in group IDEX \

3.954598 5.914478 $$
```

La  $p\text{-value} > 0.05$ , on accepte l'hypothèse  $H_0$ . Alors il n'y a pas de différence significative entre les moyennes entre **IDEX** et **beneficiaires**.

#### A.4 Test d'indépendance entre le le domaine de recherche et le programme d'investissement.

Donc d'après les résultats obtenus dans l'analyses bivarié pour ce croisement, il ne semble pas émerger de tendance de dépendance entre **ideX** et le **secteur\_discipline**. On va supposer alors les hypothèses  $H_1$ =indépendance entre les variables ( $p\text{-value} > \text{seuil}$ ) et  $H_0$ =dépendance entre les variables ( $p\text{-value} < \text{seuil}$ ) et nous allons faire le teste du khi-deux.

```
$$ chisq.test(eff_idex_secteur) \
Pearson's Chi-squared test\
data: eff_idex_secteur\
X-squared = 29.396, df = 4, p-value = 0.000006495 $$
```

Au niveau de significativité de 5 %, on conclut qu'il existe un lien de dépendance significative entre deux les variables.

Les principales contributions à l'inertie sont dues aux couples (IDEX, Santé),(IDEX, Droit et sciences économiques) respectivement sur-représentées et (IDEX, Sciences) sous-représentée par rapport à la situation théorique de dépendance. (**voir figure 15**).

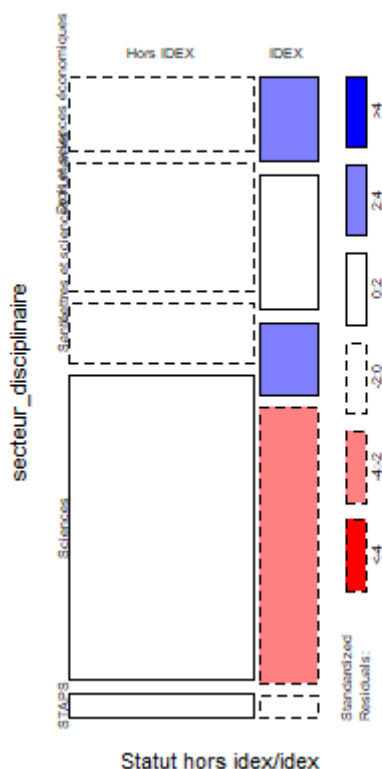


FIGURE 15 – Visualisation des résidus associés entre le programme d'investissement et le domaine de recherche.

## A.5 Test d'indépendance entre le programme d'investissement et la position du laboratoire de recherche à l'établissement.

### A.5.1 Profil ligne

L'étude du profil ligne de ce croisement nous montre que parmi les **Hors IDEX** il y a 13.6 % qui sont en **Hors PRES** et 86.4 % en **PRES**. Et parmi ceux dans **idex** il y a 20.6 % qui sont en **Hors PRES** et 79.4 % en **PRES**. ( voir figure 16 ).

### A.5.2 Profil colonne

Ici, on voit que pour la variable **Hors PRES** 67 % sont **Hors IDEX** alors que les 33 % sont **IDEX**, et pour la variable **PRES** 77 % sont **Hors IDEX** alors que les 23 % sont **IDEX** (voir figure 17)

D'après ces analyses, il ne semble pas émerger de tendance de dépendance entre le programme d'investissement et la position du laboratoire de recherche à l'établissement. On va supposer alors les hypothèses  $H_1$ =indépendance entre les variables ( $p\_value > \text{seuil}$ ) et  $H_0$ =dépendance entre les variables ( $p\_value < \text{seuil}$ ) et nous allons faire le test du khi-deux.

```
## chisq.test(eff_idex_dist) \
Pearson's Chi-squared test \
data: eff_idex_dist\
```

X-squared = 78.834, df = 1, p-value = 0.00000000000000022 ##

Au niveau de significativité de 5 %, on conclut qu'il existe un lien de dépendance significative entre les variables. Alors les principales contributions à l'inertie sont dues aux couples (HORS IDEX, Hors PRES), (IDEX, PRES) respectivement sous-représentées et (IDEX, HORS PRES) sur-représentées par rapport à la situation théorique de dépendance.(voir figure 18)

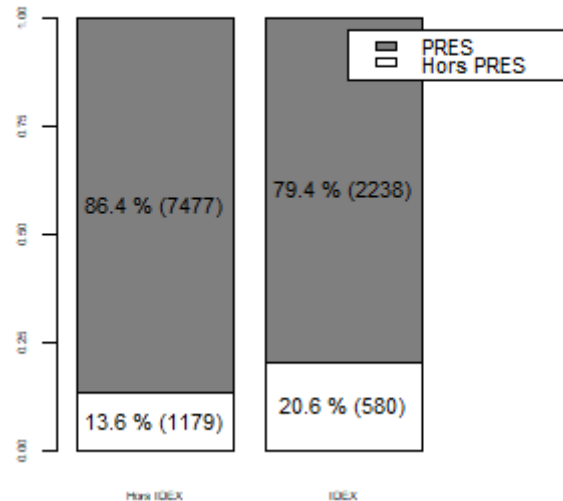


FIGURE 16 – Effectifs du profil ligne représentée des variables iDEX et la position du laboratoire de recherche à l'établissement.

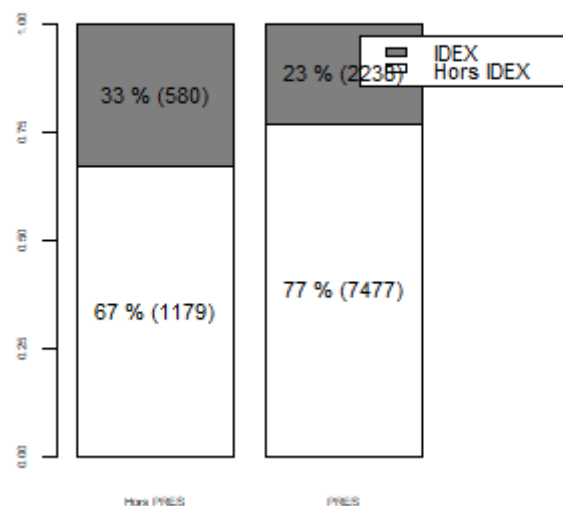


FIGURE 17 – Effectifs du profil colonne représentée des variables le programme d'investissement et la position du laboratoire de recherche à l'établissement

## A.6 Test d'indépendance entre le sexe et le nombre de bénéficiaires.

En croisant ces deux variables, nous constatons qu'il y a plus de bénéficiaires d' **Hommes** que de **femme**, et comme la variable **beneficiaires** est quantitative et **sexe** est qualitative à deux modalités, nous avons donc effectué le test de student.

```

$$ F test to compare two variances \
data:/quad beneficiaires by sexe \
F = 0.085102, num df = 4251, denom df = 7221, p-value < 0.00000000000000022 \
alternative hypothesis: true ratio of variances is not equal to 1 \
95 percent confidence interval: \
0.08067692 0.08980435 \
    
```

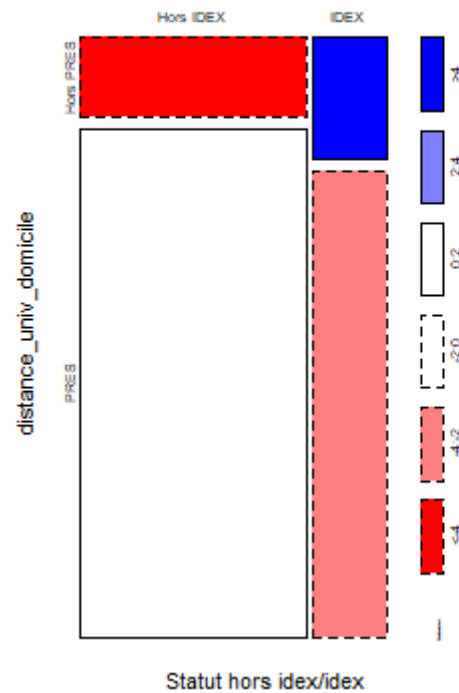


FIGURE 18 – Visualisation des résidus associés entre le programme d'investissement et la position du laboratoire de recherche à l'établissement

```
sample  estimates: \
ratio  of  variances \
```

0.08510234 \$\$

On constate une différence notable de variances au sein des sous-populations et la  $p\text{-value} < 0.05$ , on conclut qu'il n'y a pas une différence significative de variance entre les deux populations.

```
$$ Welch  Two  Sample  t-test \
```

```
data:/quad beneficiaires  by  sexe \
```

```
t = -28.235, df = 9135.2, p-value = 1 \
```

```
alternative hypothesis: true difference in means is greater than 0 \
```

```
95 percent confidence interval: \
```

```
-3.36289 Inf \
```

```
sample  estimates: \
```

```
mean in group Femmes  mean in group Hommes \
```

2.435795

5.613542 \$\$

Puisque la  $p\text{-value} > 0,05$  alors on accepte l'hypothèse  $H_0$ , et il n'y a pas de différence significative entre les moyennes.

## B Bibliographie

1. Hamon Agnès et Jégou Nicolas *Presses universitaires de Rennes, 2008*. “Statistique descriptive.”
2. Husson François et Lê, Sébastien *Pratique de la statistique* “Analyse de données avec R.”
3. Les étapes statistiques Site
4. Ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation Site
5. Les bénéficiaires de la prime d'excellence scientifique. Site .