



**MINISTERE DE
L'ENSEIGNEMENT SUPERIEUR,
DE LA RECHERCHE ET DE
L'INNOVATION**

--- ----

**UNIVERSITE JOSEPH KI-
ZERBO**

--- ----

**INSTITUT SUPERIEUR DES
SCIENCES DE LA POPULATION**

LICENCE PROFESSIONNELLE EN ANALYSE STATISTIQUE

PRATIQUE PROJET D'ANALYSE DE SURVIE

Membres du groupe 8:

OUATTARA Ousmane

SANKARA Saïdou

SAWADOGO Yacouba

ENSEIGNANT :

Dr Fabrice YAMEOGO

Table de matière

Introduction	4
I. LES VARIABLES SUSCEPTIBLES D'INFLUENCER LE TEMPS MIS POUR AVOIR LE 1ER EMPLOI	4
1. Le type de formation:	4
2. Le niveau de la formation.....	4
3. Le sexe	4
4. La période post-formation	4
5. Experience professionnelle antérieure	5
6. Lieu de résidence.....	5
II. Importation des base de données.....	5
III. Fusion des deux bases	5
IV. Selection des variables d'études	5
V. Traitement des cases vide	6
VI. ANALYSE DESCRIPTIVE DES VARIABLES RETENUES POUR L'ANALYSE.....	6
1. Tabulation de la variable phdip9: Plus haut diplôme en 9 postes	6
2. Tabulation de la variable SUPER: Sortant du Supérieur	7
3. Tabulation de la variable Q1: Sexe de l'enquêté (1:homme , 2:femme)	7
4. Tabulation de la variable AP0: Terminé un contrat d'apprentissage entre 2003 et 2024	8
5. Tabulation de la variable AP2D: Déjà travaillé ou fait un stage dans cette entreprise.....	8
6. Tabulation de la variable AP6: Dans l'entreprise, avez-vous mis en application ce que vous appreniez au centre de formation	9
7. Tabulation de la variable AP18: Contrat d'apprentissage antérieur	9
8. Tabulation de la variable ep15: Comment vous aviez su qu'il y avait une embauche possible dans cette entreprise	10
9. Tabulation de la variable stat_emb: Contrat de travail lors de l'entrée dans cette séquence entreprise	11
10. Tabulation de la variable stat_fin: Dernier contrat de travail dans cette séquence entreprise	11
VII. Création de la variables EDI.....	12
VIII. MODELISATION DE LA DUREE AVANT L'OBTENTION DU PREMIER EDI	12
1. Calculer de la durée avant l'obtention du premier EDI à la sortie du système éducatif	12
2. Modélisation de la durée avant l'obtention du premier EDI à la sortie du système éducatif	12
IX. ANALYSE DE LA FONCTION DE SURVIE	14
1. Ajustement du modèle de Kaplan-Meier par genre.....	15
2. Ajustement du modèle de Kaplan-Meier par plus haut diplôme.....	17

3. Ajustement du modèle de Kaplan-Meier par terminer un contrat d'apprentissage entre 2003 et 2004.....	18
4. Ajustement du modèle de Kaplan-Meier par sorti du supérieur	20
5. Ajustement du modèle de Kaplan-Meier par genre et sorti du supérieur.....	21
X. REGRESSION DE COX	22
Conclusion.....	27

Introduction

L'accès à l'emploi permanent pour les jeunes diplômés est un sujet d'une importance capitale dans le contexte actuel de compétitivité accrue sur le marché du travail. La présente étude vise à analyser les facteurs déterminants le délai d'obtention du premier emploi durable (EDI) pour les individus ayant quitté le système éducatif, en s'appuyant sur les données de l'enquête générationnelle CEREQ 2004. L'objectif principal est de comprendre comment diverses variables telles que le type et le niveau de formation, le sexe, l'expérience professionnelle antérieure, et le lieu de résidence influencent ce délai. Cette analyse s'appuiera sur des méthodes de survie afin de modéliser le temps jusqu'à l'obtention de l'EDI et d'identifier les déterminants clés qui peuvent orienter les politiques publiques et les stratégies d'insertion professionnelle.

I. LES VARIABLES SUSCEPTIBLES D'INFLUENCER LE TEMPS MIS POUR AVOIR LE 1ER EMPLOI

Les variables susceptibles d'influencer le temps mis pour avoir le premier emploi juste après la sortie de l'école :

1. Le type de formation:

Selon le type de formation, le temps mis pour avoir le premier emploi peut varier. Hypothèse : Les personnes ayant suivi une formation technique ou professionnelle auraient un accès plus rapide au premier emploi que les formations générales.

2. Le niveau de la formation

En fonction du niveau de la formation des personnes, le temps mis pour avoir le premier emploi peut varier. Hypothèse : Les formations de haut niveau comme niveau supérieur pourraient offrir un accès plus rapide à un premier emploi.

3. Le sexe

Les hommes et les femmes peuvent avoir des temps d'accès différents en raison de diverses contraintes socio-économiques. Hypothèse : Les hommes obtiennent un EDI plus rapidement que les femmes.

4. La période post-formation

Selon la période où la personne finit sa formation, le temps mis pour avoir son premier emploi peut varier. En période de récession, la demande pour de nouveaux travailleurs diminue, allongeant ainsi le temps nécessaire pour trouver un emploi. Hypothèse : En période de demande forte en main-d'œuvre, le temps mis pour avoir un premier emploi pourrait être réduit.

5. Experience professionnelle antérieure

Avoir une expérience professionnelle (stage, apprentissage, etc.) peut réduire le temps d'accès à un premier emploi. Hypothèse : Les individus ayant une expérience professionnelle antérieure trouveraient un EDI plus rapidement.

6. Lieu de résidence

Les régions avec un marché du travail dynamique peuvent offrir des opportunités d'avoir un premier emploi plus rapidement. Hypothèse : Les résidents des régions urbaines accèdent plus rapidement à un premier emploi que ceux des zones rurales.

II. Importation des base de données

```
library(haven)
individus <- read_sas("individus.sas7bdat", NULL)
emploi <- read_sas("emploi.sas7bdat", NULL)
```

III. Fusion des deux bases

Nous faisons la jointure des deux bases (**emploi** et **individus**) en ne conservant que les individus avec des identifiants présents dans les deux bases.

```
library(dplyr)

##
## Attachement du package : 'dplyr'

## Les objets suivants sont masqués depuis 'package:stats':
##
##      filter, lag

## Les objets suivants sont masqués depuis 'package:base':
##
##      intersect, setdiff, setequal, union

database <- inner_join(individus, emploi, by = "IDENT")
```

IV. Selection des variables d'études

phdip9 : Plus haut diplôme en 9 postes

SUPER : Sortant du Supérieur

Q1 : Sexe de l'enquêté (1:homme , 2:femme)

AP0 : Terminé un contrat d'apprentissage entre 2003 et 2024

AP2D : Déjà travaillé ou fait un stage dans cette entreprise

AP6 : Dans l'entreprise, avez-vous mis en application ce que vous appreniez au centre de formation

AP18 : Contrat d'apprentissage antérieur

ep15 : Comment vous aviez su qu'il y avait une embauche possible dans cette entreprise

DATDN : Date de début de la séquence sous forme d'indice

GEO1 : Recherche d'emploi

stat_emb: Contrat de travail lors de l'entrée dans cette séquence entreprise

stat_fin: Dernier contrat de travail dans cette séquence entreprise

idnc: Date du changement de contrat de travail

SITDE: Situation de l'enquêté à la date de l'enquête

V. Traitement des cases vides

Nous avons jugé nécessaire de traiter les cases vides et non supprimer les observations qui les contiennent parce qu'on a remarqué dans le dictionnaire des variables qu'il y avait des questions qui ne concernaient pas certains individus. Supprimer donc ces valeurs pourrait faire disparaître certains groupes d'individus.

```
library(dplyr)
library(forcats)

# Création d'une copie de database pour conserver l'original
base <- database

# Convertir les colonnes de caractère en facteur, sauf "IDENT"
base <- base %>%
  mutate(across(where(is.character) & !one_of("IDENT"), as.factor))

# Boucle pour recoder chaque colonne de facteur
for (col in names(base)) {
  # Vérifiez si la colonne est factorielle (catégorielle)
  if (is.factor(base[[col]])) {
    # Remplace les niveaux vides par NA
    levels(base[[col]])[levels(base[[col]]) == ""] <- NA
  }
}
```

VI. ANALYSE DESCRIPTIVE DES VARIABLES RETENUES POUR L'ANALYSE

1. Tabulation de la variable phdip9: Plus haut diplôme en 9 postes

```
round(prop.table(table(base$phdip9, base$sitde), margin=1),4)
```

	Chômage	Emploi	Formation	Inactivité	Reprise d'étu
des					
## Bac	0.1088	0.7948	0.0304	0.0211	0.0
449					
## BTS-DUT-Santé-social	0.0560	0.9060	0.0088	0.0125	0.0
167					
## CAP-BEP-MC	0.1265	0.8178	0.0163	0.0264	0.0
129					
## DEA-DESS-Master-M2	0.0504	0.9233	0.0066	0.0093	0.0
104					

##	Deug	0.0836	0.7826	0.0394	0.0191	0.0
753						
##	Doctorat	0.0641	0.9218	0.0042	0.0071	0.0
029						
##	Licence L3	0.0666	0.8564	0.0197	0.0236	0.0
337						
##	Maîtrise M1, MST...	0.0872	0.8488	0.0139	0.0241	0.0
260						
##	Non diplômé	0.1963	0.7287	0.0233	0.0304	0.0
213						

Ce tableau est le croisement de la variable “Plus haut diplôme” et la “situation de la personne pendant l’enquête”. Pour tout type de diplôme, ceux qui ont un emploi sont les plus nombreux avec respectivement 79.48% pour ceux qui ont le bac, 90.60% pour BTS-DUT-Santé-social, 81.78% pour CAP-BEP-MC, 92.23% pour DEA-DESS-Master-M2, 78.26% pour Deug, 92.18% pour Doctorat, 85.64% pour Licence L3, 84.88% pour Maîtrise M1, MST et 72.87% pour les non diplômés.

2. Tabulation de la variable SUPER: Sortant du Supérieur

```
table(base$SUPER)

##
##      Atteint niveau supérieur Pas atteint niveau supérieur
##              39122                  30860

round(prop.table(table(base$SUPER, base$sitde), margin=1),4)

##
##              Chômage Emploi Formation Inactivité
##      Atteint niveau supérieur      0.0717 0.8676      0.0162      0.0177
##      Pas atteint niveau supérieur 0.1357 0.7980      0.0207      0.0231
##
##              Reprise d'études
##      Atteint niveau supérieur      0.0268
##      Pas atteint niveau supérieur      0.0225
```

Les personnes qui ont atteint un niveau supérieur sont de 39122 contre 30860 qui n'ont pas atteint un niveau supérieur. Ces statistiques nous montrent que ceux qui ont atteint le niveau supérieur ont une proportion d'accès à l'emploi plus élevée que ceux qui n'ont pas atteint niveau supérieur, soit 86.76% contre 79.80%.

3. Tabulation de la variable Q1: Sexe de l'enquête (1:homme , 2:femme)

```
table(base$Q1)

##
##      femme homme
##      33414 36568

round(prop.table(table(base$Q1, base$sitde), margin=1),4)

##
##              Chômage Emploi Formation Inactivité Reprise d'études
```

##	femme	0.0965	0.8257	0.0227	0.0245	0.0306
##	homme	0.1030	0.8472	0.0141	0.0160	0.0197

Les femmes sont 33414 tandis que les hommes sont de 36568. La différence de proportion d'accès à l'emploi entre les hommes et les femmes n'est pas très remarquable soit respectivement 84.72% et 82.57%.

4. Tabulation de la variable AP0: Terminé un contrat d'apprentissage entre 2003 et 2024

```
table(base$AP0)

##
## contrat d'apprentissage non terminé      contrat d'apprentissage terminé
##                                564                                10945

round(prop.table(table(base$AP0, base$sitde), margin=1),4)

##
##                                Chômage Emploi Formation Inactivit
é
## contrat d'apprentissage non terminé  0.0851 0.8582    0.0142    0.019
5
## contrat d'apprentissage terminé      0.0867 0.8831    0.0081    0.015
3
##
##                                Reprise d'études
## contrat d'apprentissage non terminé      0.0230
## contrat d'apprentissage terminé          0.0068
```

Les personnes n'ayant pas terminé leur contrat d'apprentissage sont de 564 et celles ayant terminé leur contrat d'apprentissage sont de 10945. La différence de la proportion d'accès à l'emploi entre les personnes qui ont eu un contrat d'apprentissage non terminé et celles qui ont eu un contrat d'apprentissage terminé n'est pas très grande soit respectivement 85.82% et 88.31%.

5. Tabulation de la variable AP2D: Déjà travaillé ou fait un stage dans cette entreprise

```
table(base$AP2D)

##
## ni travail ni stage      travail ou stage
##                7280                3665

round(prop.table(table(base$AP2D, base$sitde), margin=1),4)

##
##                                Chômage Emploi Formation Inactivité Reprise d'étud
es
## ni travail ni stage  0.0870 0.8765    0.0098    0.0187                0.00
81
## travail ou stage     0.0862 0.8963    0.0049    0.0085                0.00
41
```


Les personnes qui n'ont ni travaillé, ni fait de stage dans l'entreprise sont de 59037 tandis que celles qui ont travaillé ou fait de stage sont de 3665. La différence de la proportion d'accès à l'emploi entre les personnes qui n'ont ni travaillé ni fait de stage dans l'entreprise et qui ont travaillé ou fait de stage dans l'entreprise, n'est pas très grande soit respectivement 87.65% et 89.63%.

6. Tabulation de la variable AP6: Dans l'entreprise, avez-vous mis en application ce que vous appreniez au centre de formation

```
table(base$AP6)

##
##      Non, jamais      Oui, parfois      Oui, rarement      Oui, souvent
##              994              2567              898              6486

round(prop.table(table(base$AP6, base$sitde), margin=1),4)

##
##              Chômage      Emploi      Formation      Inactivité      Reprise d'études
##      Non, jamais      0.1127      0.8561      0.0020      0.0221      0.0070
##      Oui, parfois      0.0830      0.8929      0.0074      0.0117      0.0051
##      Oui, rarement      0.1392      0.8330      0.0067      0.0156      0.0056
##      Oui, souvent      0.0769      0.8904      0.0096      0.0156      0.0076
```

Les personnes qui n'ont jamais mis en application ce qu'elles ont appris sont de 994, celles qui ont parfois appliqué sont de 2567, celles qui ont rarement appliqué sont de 898, celles qui ont souvent appliqué sont de 6486. La situation des personnes selon s'elles qui ont appliqué leurs connaissances au centre de formation diffère sensiblement. Celles qui n'ont jamais appliqué ont une proportion d'accès à l'emploi de 85.61%. Celles qui appliquent parfois et celles qui appliquent souvent ont toutes une proportion d'accès à l'emploi de 89%. Quant à celles qui appliquent rarement, elles ont une proportion d'accès à l'emploi de 83.30%.

7. Tabulation de la variable AP18: Contrat d'apprentissage antérieur

```
table(base$AP18)

##
##      Contrat d'apprentissage antérieur
##              2263
##      pas de Contrat d'apprentissage antérieur
##              8682

round(prop.table(table(base$AP18, base$sitde), margin=1),4)

##
##              Chômage      Emploi      Formation      Inac
##      tativité
##      Contrat d'apprentissage antérieur      0.0826      0.8860      0.0119
##      0.0163
##      pas de Contrat d'apprentissage antérieur      0.0878      0.8824      0.0071
##      0.0150
##
##              Reprise d'études
```

##	Contrat d'apprentissage antérieur	0.0031
##	pas de Contrat d'apprentissage antérieur	0.0077

La différence de la proportion d'accès à l'emploi entre les personnes qui ont eu Contrat d'apprentissage antérieur et celles qui n'ont pas eu Contrat d'apprentissage antérieur, n'est pas très grande soit respectivement 88.60% et 88.24%.

8. Tabulation de la variable ep15: Comment vous aviez su qu'il y avait une embauche possible dans cette entreprise

```
table(base$ep15)
```

```
##
##          (NE PAS CITER) Par l'intérim
##                               4818
## (NE PAS CITER) Par une candidature spontanée
##                               11938
##                               Autres
##                               6192
##                               L'APEC
##                               264
##                               Par l'ANPE
##                               4182
##                               Par une de vos relations
##                               16961
##                               Par une mission locale, une PAIO
##                               582
## Par une petite annonce (presse, internet)
##                               6585
##                               Votre établissement de formation
##                               4256
```

```
round(prop.table(table(base$ep15, base$sitde), margin=1),4)
```

```
##
##                               Chômage Emploi Formation
## (NE PAS CITER) Par l'intérim      0.1343 0.8045    0.0189
## (NE PAS CITER) Par une candidature spontanée 0.0849 0.8539    0.0190
## Autres                          0.0610 0.8929    0.0141
## L'APEC                          0.0379 0.9583    0.0000
## Par l'ANPE                      0.1148 0.8302    0.0167
## Par une de vos relations        0.1123 0.8191    0.0182
## Par une mission locale, une PAIO 0.1907 0.7371    0.0223
## Par une petite annonce (presse, internet) 0.0853 0.8685    0.0106
## Votre établissement de formation 0.0623 0.9018    0.0078
##
##                               Inactivité Reprise d'étud
es
## (NE PAS CITER) Par l'intérim      0.0224          0.01
99
## (NE PAS CITER) Par une candidature spontanée 0.0196          0.02
26
## Autres                          0.0131          0.01
89
```

##	L 'APEC	0.0038	0.00
00			
##	Par l'ANPE	0.0206	0.01
77			
##	Par une de vos relations	0.0232	0.02
74			
##	Par une mission locale, une PAIO	0.0258	0.02
41			
##	Par une petite annonce (presse, internet)	0.0146	0.02
10			
##	Votre établissement de formation	0.0101	0.01
81			

La situation des personnes selon la source d'information qu'il y avait une embauche possible dans cette entreprise diffère. Celles qui ont eu l'information à travers l'ANPE ont une proportion d'accès à l'emploi de 83.02%. Celles qui ont eu l'information par les relations ont une proportion d'accès à l'emploi de 81.91%. Celles qui ont eu l'information par une mission locale ou une PAIO ont une proportion d'accès à l'emploi de 73.71%. Celles qui ont eu l'information par une petite annonce (presse, internet) ont une proportion d'accès à l'emploi de 86.85%. Quant à celles qui ont eu l'information à travers leur établissement de formation ont une proportion d'accès à l'emploi de 90.18%.

9. Tabulation de la variable stat_emb: Contrat de travail lors de l'entrée dans cette séquence entreprise

```
table(base$stat_emb_ED1)

##
##  Autres emb_ED1
##   50854   19128

round(prop.table(table(base$stat_emb_ED1, base$site), margin=1),4)

##
##           Chômage Emploi Formation Inactivité Reprise d'études
##  Autres   0.1202 0.8042   0.0227   0.0227           0.0301
##  emb_ED1  0.0458 0.9240   0.0062   0.0131           0.0109
```

La situation des personnes à l'embauchement diffère. Celles qui ont eu un EDI ont un accès à l'emploi de 92.40% contre 80.42% pour les autres.

10. Tabulation de la variable stat_fin: Dernier contrat de travail dans cette séquence entreprise

```
table(base$stat_fin_ED1)

##
##  Autres fin_ED1
##   45394   24588

round(prop.table(table(base$stat_fin_ED1, base$site), margin=1),4)
```

```
##
##           Chômage Emploi Formation Inactivité Reprise d'études
##  Autres   0.1326 0.7839    0.0251    0.0250    0.0334
##  fin_EDI  0.0395 0.9348    0.0055    0.0109    0.0092
```

La situation des personnes au dernier contrat de travail dans cette séquence entreprise diffère. Celles qui ont eu un EDI ont un accès à l'emploi de 93.48% contre 78.39% pour les autres.

VII. Création de la variables EDI

```
base <- base %>%
  mutate(EDI = case_when(stat_emb=="03" ~ 1, stat_emb=="04" ~ 1 ,stat_fin=="
03" ~ 1,stat_fin=="04" ~ 1, TRUE ~ 0))
```

VIII. MODELISATION DE LA DUREE AVANT L'OBTENTION DU PREMIER EDI

1. Calculer de la durée avant l'obtention du premier EDI à la sortie du système éducatif

```
base <- base %>%
  mutate(id=as.numeric(id),DATDN=as.numeric(DATDN),`if`=as.numeric(`if`),idnc
=as.numeric(idnc))

library(dplyr)
base <- base %>%
  mutate(duree = case_when(
    EDI == 1 & (stat_emb != "04" & stat_emb != "03") ~ as.numeric(idnc - id
),
    EDI == 1 & (stat_fin != "04" & stat_fin != "03") ~ as.numeric(DATDN - i
d),
    TRUE ~ `if` - id
  ))
```

2. Modélisation de la durée avant l'obtention du premier EDI à la sortie du système éducatif

```
library(dplyr)
library(survival)
library(survminer)

## Le chargement a nécessité le package : ggplot2
## Le chargement a nécessité le package : ggpubr

##
## Attachement du package : 'survminer'

## L'objet suivant est masqué depuis 'package:survival':
##
##      myeloma

# Ajustement du modèle de Kaplan-Meier
km_fit <- survfit(Surv(duree, EDI) ~ 1, data = base)
```

```
# Résultats du modèle
```

```
summary(km_fit)
```

```
## Call: survfit(formula = Surv(duree, EDI) ~ 1, data = base)
```

```
##
```

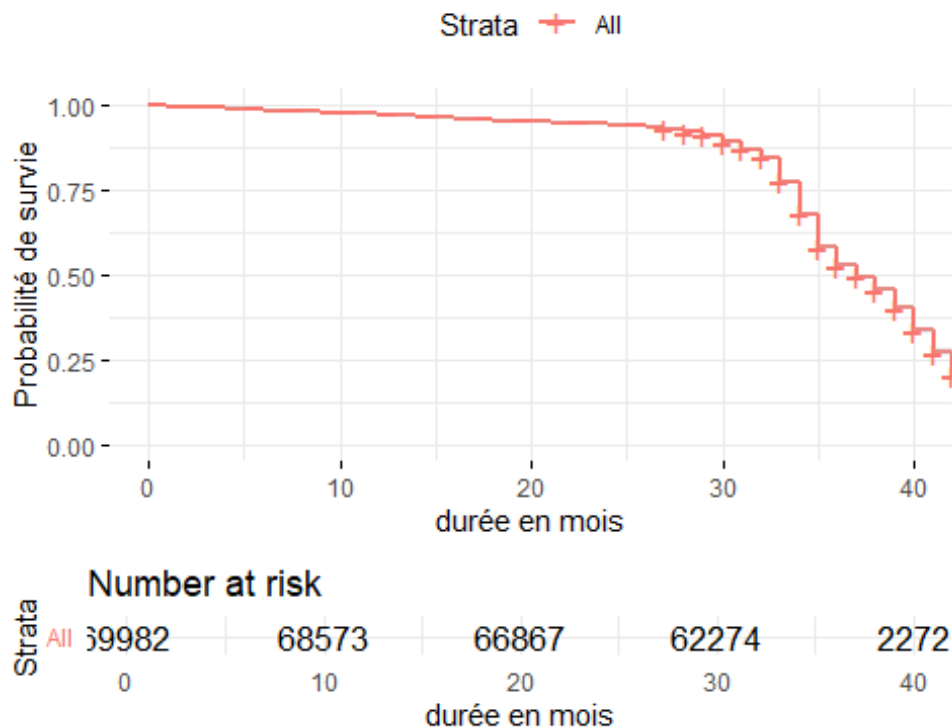
##	time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95% CI
##	1	69982	65	0.999	0.000115		0.999		0.999
##	2	69917	120	0.997	0.000194		0.997		0.998
##	3	69797	152	0.995	0.000262		0.995		0.996
##	4	69645	129	0.993	0.000307		0.993		0.994
##	5	69516	174	0.991	0.000360		0.990		0.992
##	6	69342	265	0.987	0.000427		0.986		0.988
##	7	69077	182	0.984	0.000467		0.984		0.985
##	8	68895	170	0.982	0.000502		0.981		0.983
##	9	68725	152	0.980	0.000531		0.979		0.981
##	10	68573	129	0.978	0.000554		0.977		0.979
##	11	68444	161	0.976	0.000582		0.975		0.977
##	12	68283	188	0.973	0.000612		0.972		0.974
##	13	68095	177	0.971	0.000640		0.969		0.972
##	14	67918	210	0.968	0.000670		0.966		0.969
##	15	67708	200	0.965	0.000698		0.963		0.966
##	16	67508	136	0.963	0.000716		0.961		0.964
##	17	67372	137	0.961	0.000734		0.959		0.962
##	18	67235	217	0.958	0.000761		0.956		0.959
##	19	67018	151	0.955	0.000780		0.954		0.957
##	20	66867	156	0.953	0.000798		0.952		0.955
##	21	66711	145	0.951	0.000815		0.950		0.953
##	22	66566	135	0.949	0.000830		0.948		0.951
##	23	66431	139	0.947	0.000845		0.946		0.949
##	24	66292	190	0.945	0.000865		0.943		0.946
##	25	66102	160	0.942	0.000882		0.941		0.944
##	26	65942	279	0.938	0.000910		0.937		0.940
##	27	65663	339	0.933	0.000942		0.932		0.935
##	28	65087	627	0.924	0.000999		0.922		0.926
##	29	63888	795	0.913	0.001067		0.911		0.915
##	30	62274	1432	0.892	0.001178		0.890		0.894
##	31	58990	1281	0.873	0.001270		0.870		0.875
##	32	55941	1552	0.848	0.001376		0.846		0.851
##	33	52198	4480	0.776	0.001632		0.772		0.779
##	34	35424	4268	0.682	0.001965		0.678		0.686
##	35	19185	2790	0.583	0.002415		0.578		0.588
##	36	8761	806	0.529	0.002837		0.524		0.535
##	37	5988	369	0.497	0.003129		0.491		0.503
##	38	4768	377	0.457	0.003475		0.451		0.464
##	39	3504	410	0.404	0.003947		0.396		0.412
##	40	2272	377	0.337	0.004558		0.328		0.346
##	41	1186	231	0.271	0.005336		0.261		0.282
##	42	451	107	0.207	0.006789		0.194		0.221
##	43	102	23	0.160	0.010047		0.142		0.181
##	44	25	5	0.128	0.015131		0.102		0.162

```
# Visualisation de la fonction de survie estimée
```

```
ggsurvplot(km_fit, data = base, xlab = "durée en mois", ylab = "Probabilité
```

de survie",

```
conf.int = TRUE, risk.table = TRUE, ggtheme = theme_minimal())
```



Ce graphique montre la probabilité de survie (ne pas avoir un EDI) de 69 982 individus sur une période de 40 mois. Au fil du temps, la chance d'avoir un EDI augmente progressivement et devient plus prononcée après 30 mois. À 40 mois, environ 73 % des individus ont eu un EDI. Le nombre de personnes à risque passe de 69 982 au début à 2 272 à 40 mois. En résumé, la chance d'avoir un EDI augmente régulièrement.

IX. ANALYSE DE LA FONCTION DE SURVIE

Ajustement du modèle de Kaplan-Meier

```
km_fit <- survfit(Surv(duree, EDI) ~ 1, data = base)
```

Résultats du modèle

```
summary(km_fit)
```

```
## Call: survfit(formula = Surv(duree, EDI) ~ 1, data = base)
```

```
##
```

##	time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
##	1	69982	65	0.999	0.000115	0.999	0.999
##	2	69917	120	0.997	0.000194	0.997	0.998
##	3	69797	152	0.995	0.000262	0.995	0.996
##	4	69645	129	0.993	0.000307	0.993	0.994
##	5	69516	174	0.991	0.000360	0.990	0.992
##	6	69342	265	0.987	0.000427	0.986	0.988
##	7	69077	182	0.984	0.000467	0.984	0.985
##	8	68895	170	0.982	0.000502	0.981	0.983
##	9	68725	152	0.980	0.000531	0.979	0.981
##	10	68573	129	0.978	0.000554	0.977	0.979
##	11	68444	161	0.976	0.000582	0.975	0.977
##	12	68283	188	0.973	0.000612	0.972	0.974
##	13	68095	177	0.971	0.000640	0.969	0.972

##	14	67918	210	0.968	0.000670	0.966	0.969
##	15	67708	200	0.965	0.000698	0.963	0.966
##	16	67508	136	0.963	0.000716	0.961	0.964
##	17	67372	137	0.961	0.000734	0.959	0.962
##	18	67235	217	0.958	0.000761	0.956	0.959
##	19	67018	151	0.955	0.000780	0.954	0.957
##	20	66867	156	0.953	0.000798	0.952	0.955
##	21	66711	145	0.951	0.000815	0.950	0.953
##	22	66566	135	0.949	0.000830	0.948	0.951
##	23	66431	139	0.947	0.000845	0.946	0.949
##	24	66292	190	0.945	0.000865	0.943	0.946
##	25	66102	160	0.942	0.000882	0.941	0.944
##	26	65942	279	0.938	0.000910	0.937	0.940
##	27	65663	339	0.933	0.000942	0.932	0.935
##	28	65087	627	0.924	0.000999	0.922	0.926
##	29	63888	795	0.913	0.001067	0.911	0.915
##	30	62274	1432	0.892	0.001178	0.890	0.894
##	31	58990	1281	0.873	0.001270	0.870	0.875
##	32	55941	1552	0.848	0.001376	0.846	0.851
##	33	52198	4480	0.776	0.001632	0.772	0.779
##	34	35424	4268	0.682	0.001965	0.678	0.686
##	35	19185	2790	0.583	0.002415	0.578	0.588
##	36	8761	806	0.529	0.002837	0.524	0.535
##	37	5988	369	0.497	0.003129	0.491	0.503
##	38	4768	377	0.457	0.003475	0.451	0.464
##	39	3504	410	0.404	0.003947	0.396	0.412
##	40	2272	377	0.337	0.004558	0.328	0.346
##	41	1186	231	0.271	0.005336	0.261	0.282
##	42	451	107	0.207	0.006789	0.194	0.221
##	43	102	23	0.160	0.010047	0.142	0.181
##	44	25	5	0.128	0.015131	0.102	0.162

Interprétation : Le modèle par approche non paramétrique montre une diminution progressive de la fonction de survie avec le temps avant l'obtention d'un EDI. Au début (temps 1), presque tous les individus n'ont pas encore obtenu un EDI (survie de 0.999). À temps 10, 97.8% des individus n'ont pas encore obtenu un EDI, et à temps 30, cette proportion est de 89.2%. La tendance se poursuit, et à temps 40, seulement 33.7% des individus n'ont pas encore obtenu un EDI. Finalement, à temps 44, la survie est de 0.128, indiquant que seulement 12.8% des individus n'ont pas encore obtenu un EDI. Ces résultats montrent que la probabilité de ne pas avoir obtenu un EDI diminue avec le temps, confirmant que de plus en plus d'individus obtiennent un EDI au fil du temps.

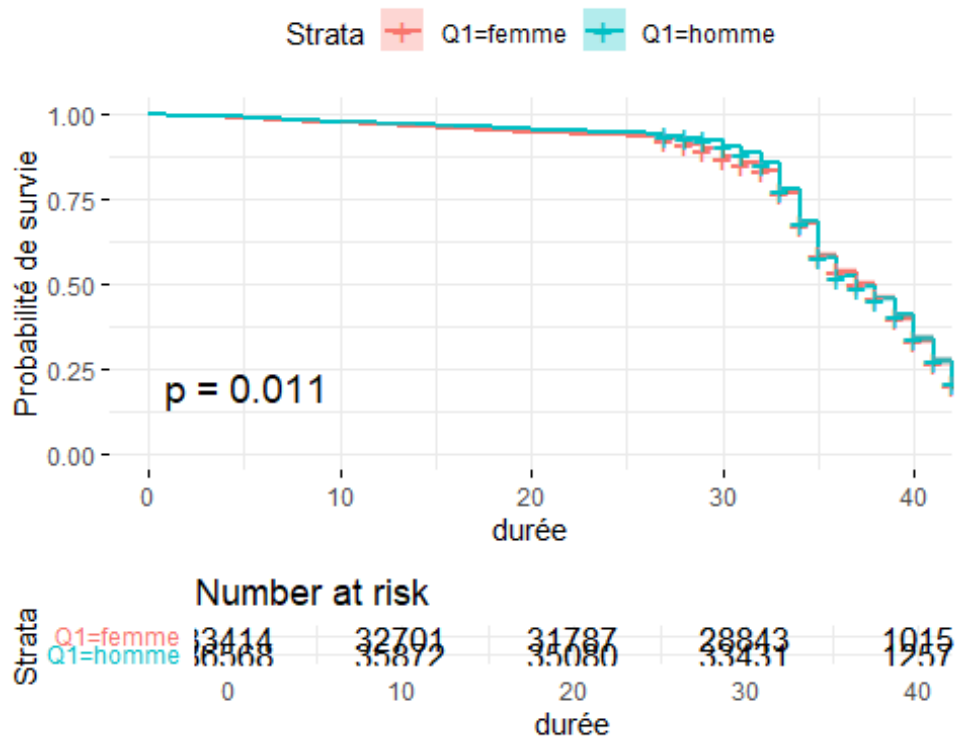
1. Ajustement du modèle de Kaplan-Meier par genre

```
km_fit_sexe <- survfit(Surv(duree, EDI) ~ Q1, data = base)

# Résultats du modèle
summary(km_fit_sexe)

# Visualisation de la fonction de survie estimée par genre
ggsurvplot(km_fit_sexe, data = base, xlab = "durée", ylab = "Probabilité de survie",
```

```
conf.int = TRUE, risk.table = TRUE, pval = TRUE, ggtheme = theme_minimal())
```



Interprétation : Toute chose égale par ailleurs, le modèle de Kaplan-Meier montre que la probabilité de ne pas avoir obtenu un EDI diminue avec le temps pour les deux sexes. Au temps 1, presque tous les individus (hommes et femmes) n'ont pas encore obtenu un CDI. À temps 30, 87.4% des femmes et 90.8% des hommes n'ont pas encore obtenu un EDI. À temps 40, 33.4% des femmes et 34% des hommes n'ont pas encore obtenu un EDI. Finalement, à temps 44, seulement 9.63% des femmes et 13.9% des hommes n'ont pas encore obtenu un EDI, montrant que les femmes obtiennent un CDI légèrement plus rapidement que les hommes.

Les tests de log-rank fournissent une statistique de test et une valeur p pour évaluer si les différences entre les courbes de survie sont statistiquement significatives. Une valeur p inférieure à 0,05 indique généralement que les différences entre les groupes sont significatives.

```
# Test de Log-rank pour comparer Les courbes de survie par genre
survdif(Surv(duree, EDI) ~ Q1, data = base)
```

```
## Call:
## survdif(formula = Surv(duree, EDI) ~ Q1, data = base)
##
##              N Observed Expected (O-E)^2/E (O-E)^2/V
## Q1=femme 33414      11497      11305        3.25        6.49
## Q1=homme 36568      13091      13283        2.77        6.49
##
## Chisq= 6.5 on 1 degrees of freedom, p= 0.01
```

Interprétation : La statistique du chi-carré obtenue est de 6.5, ce qui mesure la différence globale entre les courbes de survie des deux groupes. La p-value associée est

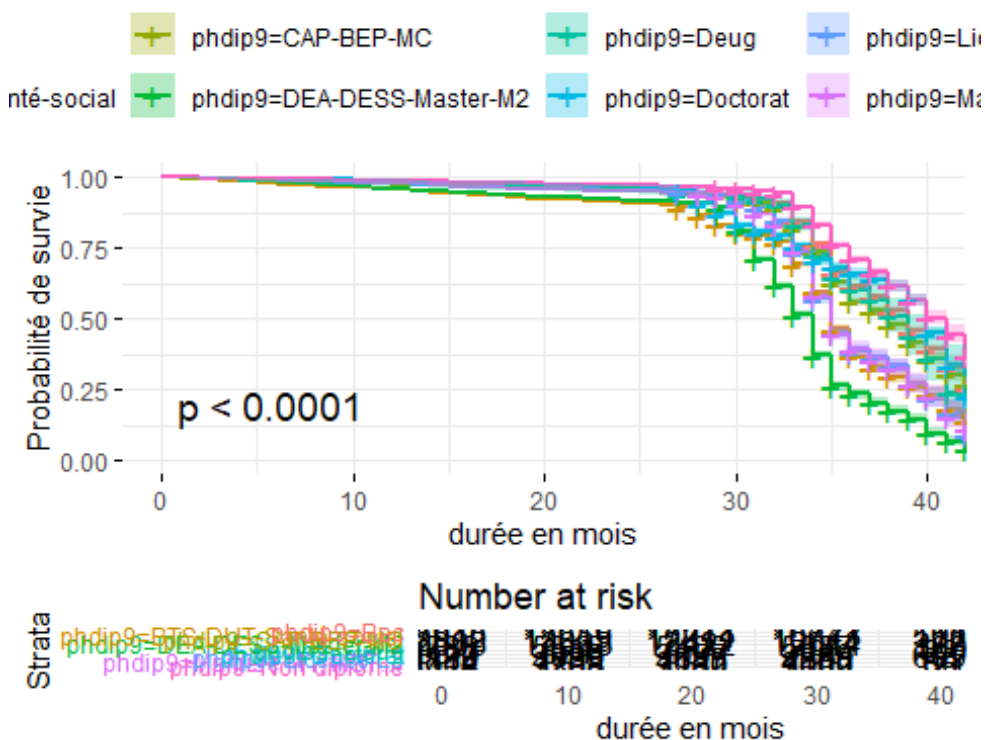
de 0.01, inférieure au seuil conventionnel de 0.05, ce qui indique une différence statistiquement significative. Cette différence suggère que les courbes de survie des hommes et des femmes ne sont pas identiques. En d'autres termes, il existe une différence significative dans la durée avant l'obtention d'un EDI entre les deux genres. Les femmes obtiennent un EDI plus rapidement que les hommes, comme le confirment les différences observées entre les courbes de survie.

2. Ajustement du modèle de Kaplan-Meier par plus haut diplôme

```
km_fit_haut_diplome <- survfit(Surv(duree, EDI) ~ phdip9, data = base)

# Résultats du modèle
summary(km_fit_haut_diplome)

# Visualisation de la fonction de survie estimée par niveau de sortie
ggsurvplot(km_fit_haut_diplome, data = base, xlab = "durée en mois", ylab = "Probabilité de survie",
            conf.int = TRUE, risk.table = TRUE, pval = TRUE, ggtheme = theme_minimal())
```



Interprétation : En supposant toute chose égale par ailleurs, les courbes de survie montrent que les titulaires de DEA-DESS-Master-M2 ont la meilleure probabilité d'obtenir un EDI rapidement, bien que cette probabilité diminue rapidement avec le temps. Les personnes ayant un BTS-DUT-Santé-social bénéficient d'une probabilité plus stable mais plus faible par rapport au DEA-DESS-Master-M2. Les titulaires de CAP-BEP-MC commencent avec une probabilité élevée, mais celle-ci diminue de manière constante et reste inférieure à celle des diplômés du DEA-DESS-Master-M2. Enfin, ceux avec un Bac font face à une diminution rapide de leur probabilité de survie, indiquant

des difficultés accrues pour obtenir un EDI. Ces résultats soulignent l'impact significatif du niveau d'éducation sur l'accès à un EDI.

Test de Log-rank pour comparer Les courbes de niveau de sortie

```
survdif(Surv(duree, EDI) ~ phdip9, data = base)
```

```
## Call:
```

```
## survdif(formula = Surv(duree, EDI) ~ phdip9, data = base)
```

```
##
```

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
## phdip9=Bac	17945	5162	6896	436.2	654.6
## phdip9=BTS-DUT-Santé-social	13812	5532	3832	754.0	965.0
## phdip9=CAP-BEP-MC	12906	4081	4848	121.3	163.1
## phdip9=DEA-DESS-Master-M2	5788	3163	1375	2327.0	2626.6
## phdip9=Deug	837	253	312	11.0	12.0
## phdip9=Doctorat	3120	1246	1480	37.0	45.6
## phdip9=Licence L3	4834	2095	1588	162.1	185.8
## phdip9=Maîtrise M1, MST...	3314	1340	1030	93.4	104.3
## phdip9=Non diplômé	7426	1716	3228	708.2	887.0

```
## Chisq= 5055 on 8 degrees of freedom, p= <2e-16
```

Interprétations : Les résultats montrent des différences significatives entre les courbes de survie des différents niveaux d'éducation en ce qui concerne l'accès à un EDI. En d'autres termes, le temps nécessaire pour obtenir un EDI varie de manière significative selon le niveau d'éducation. Le test de log-rank confirme que ces différences ne sont pas dues au hasard et que des interventions spécifiques pour chaque niveau éducatif pourraient être justifiées pour améliorer l'accès à des emplois permanents (EDI).

3. Ajustement du modèle de Kaplan-Meier par terminer un contrat d'apprentissage entre 2003 et 2004

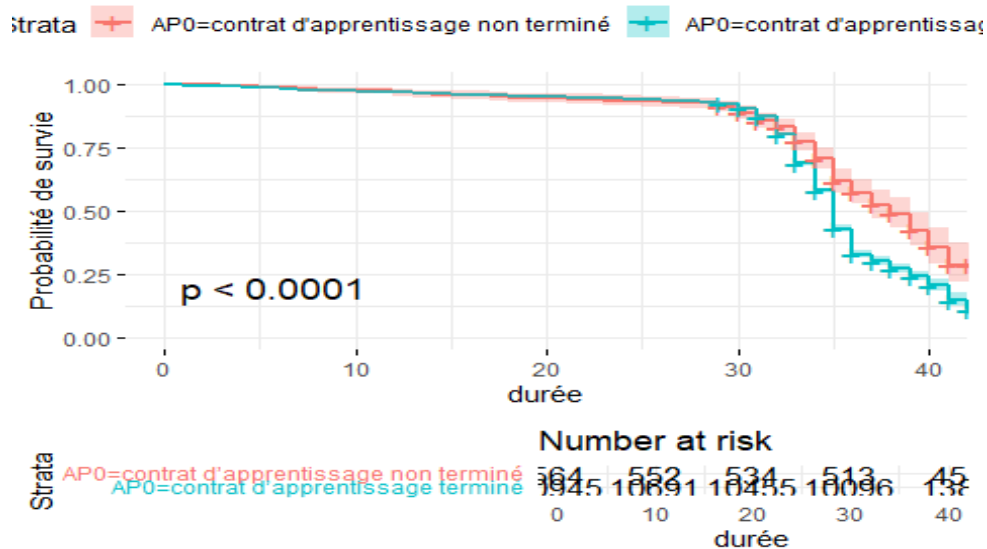
```
km_fit_AP0 <- survfit(Surv(duree, EDI) ~ AP0, data = base)
```

Résultats du modèle

```
summary(km_fit_AP0)
```

Visualisation de la fonction de survie estimée par terminer un contrat d'apprentissage entre 2003 et 2004

```
ggsurvplot(km_fit_AP0, data = base, xlab = "durée", ylab = "Probabilité de survie",  
            conf.int = TRUE, risk.table = TRUE, pval = TRUE, ggtheme = theme_minimal())
```



Interprétations : les résultats du modèle de survie montrent que les individus ayant terminé leur contrat d'apprentissage ont une probabilité de survie beaucoup plus élevée, En supposant tous les autres facteurs constants, donc ils gagnent une meilleure chance d'obtenir un EDI, comparativement à ceux dont le contrat est resté inachevé. Pour les contrats terminés, la probabilité de survie reste élevée et décroît rapidement, tandis que pour les contrats non terminés, elle diminue lentement. Cela souligne l'importance de terminer les contrats d'apprentissage pour améliorer les chances d'accès à un emploi stable

Test de Log-rank pour comparer Les courbes de terminer un contrat d'apprentissage entre 2003 et 2004

```
survdif(Surv(duree, EDI) ~ AP0, data = base)
```

```
## Call:
```

```
## survdif(formula = Surv(duree, EDI) ~ AP0, data = base)
```

```
##
```

```
## n=11509, 58473 observations effacées parce que manquantes.
```

```
##
```

```
##
```

```
##
```

```
## AP0=contrat d'apprentissage non terminé 564 227 343 39.4
```

```
6
```

```
## AP0=contrat d'apprentissage terminé 10945 4863 4747 2.8
```

```
5
```

```
## (O-E)^2/V
```

```
## AP0=contrat d'apprentissage non terminé 50
```

```
## AP0=contrat d'apprentissage terminé 50
```

```
##
```

```
## Chisq= 50 on 1 degrees of freedom, p= 2e-12
```

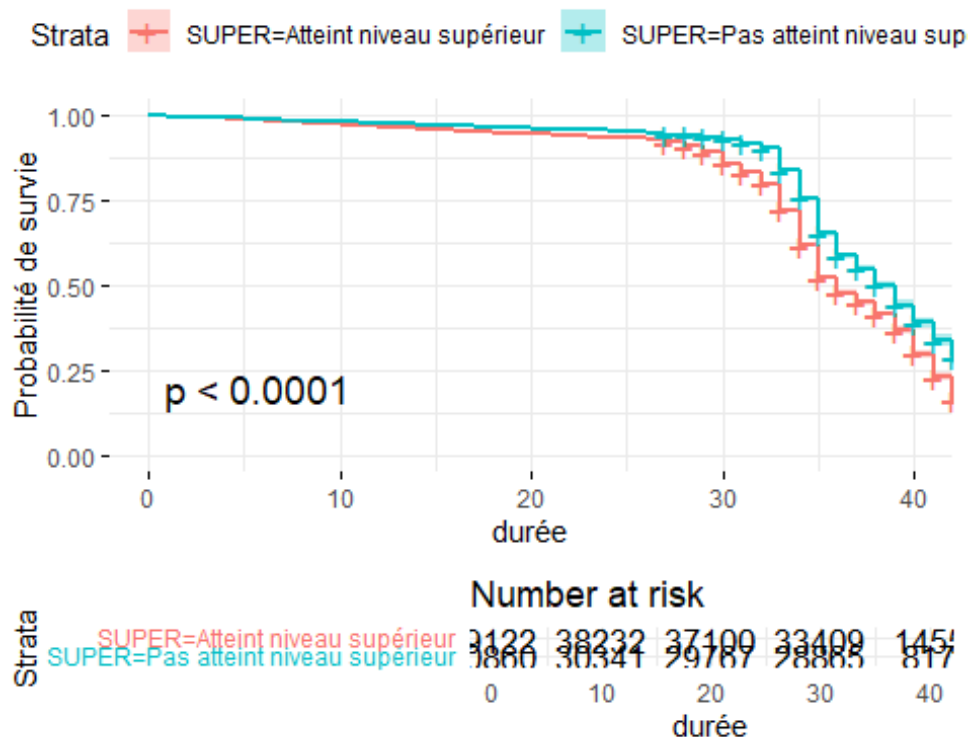
Interprétations : Le test de log-rank montre une différence significative entre les courbes de survie des individus ayant terminé leur contrat d'apprentissage et ceux dont le contrat est resté inachevé. Les personnes ayant terminé leur contrat ont une probabilité beaucoup plus élevée d'obtenir un emploi stable comparé à celles qui n'ont pas terminé leur contrat. Les courbes de survie indiquent que les chances d'obtenir un EDI diminuent plus lentement pour ceux ayant complété leur contrat d'apprentissage.

4. Ajustement du modèle de Kaplan-Meier par sorti du supérieur

```
km_fit_super <- survfit(Surv(duree, EDI) ~ SUPER, data = base)

# Résultats du modèle
summary(km_fit_super)

# Visualisation de la fonction de survie estimée par zone géographique
ggsurvplot(km_fit_super, data = base, xlab = "durée", ylab = "Probabilité d
e survie",
            conf.int = TRUE, risk.table = TRUE, pval = TRUE, ggtheme = theme
_minimal())
```



Interprétations : Les individus ayant atteint un niveau supérieur ont une probabilité plus élevée d'obtenir un EDI par rapport à ceux qui n'ont pas atteint ce niveau toute chose égale par ailleurs. Cette probabilité reste supérieure au cours du temps pour ceux ayant atteint un niveau supérieur, tandis que celle des autres diminue plus rapidement. En somme, atteindre un niveau supérieur est un facteur favorable pour obtenir et maintenir un EDI.

```
# Test de Log-rank pour comparer Les courbes de sorti de supérieur
survdif(Surv(duree, EDI) ~ SUPER, data = base)

## Call:
## survdif(formula = Surv(duree, EDI) ~ SUPER, data = base)
##
##
```

	N	Observed	Expected	(O-E)^2/E
SUPER=Atteint niveau supérieur	39122	15624	12906	572
SUPER=Pas atteint niveau supérieur	30860	8964	11682	632

```
1305
##
## Chisq= 1305 on 1 degrees of freedom, p= <2e-16
```

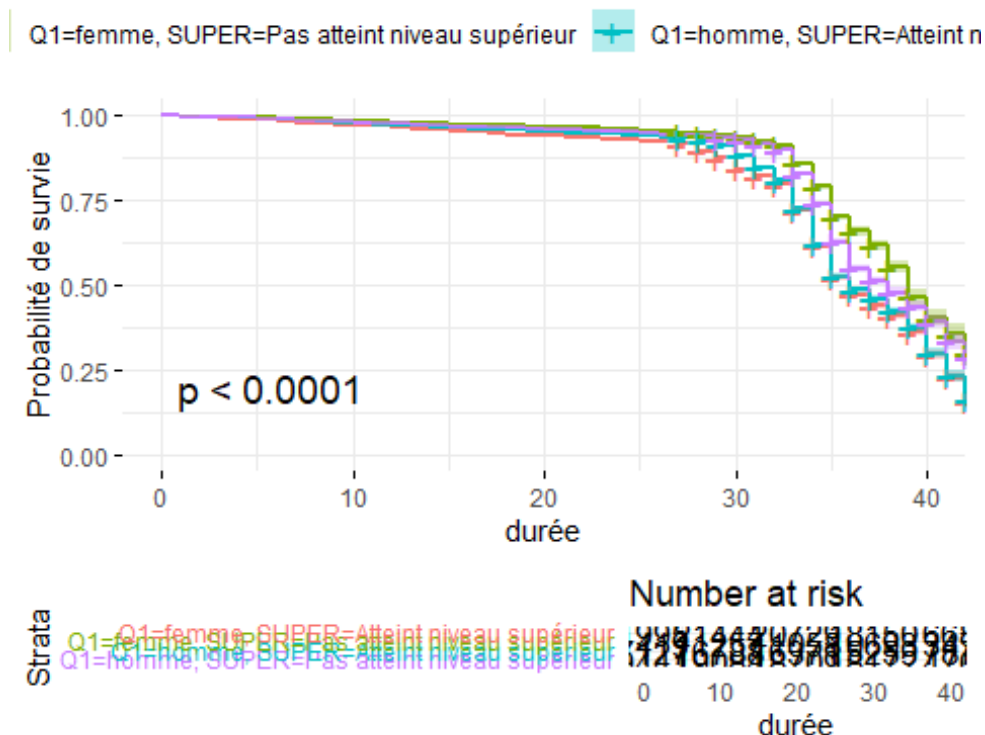
Interprétations : Les résultats montrent une différence statistiquement significative entre les deux groupes (Chi-carré = 1305, $p < 2e-16$). Les individus ayant atteint un niveau supérieur ont, en moyenne, un taux élevé de maintien d'un EDI comparé à ceux qui ne l'ont pas atteint. Cela confirme que l'atteinte d'un niveau supérieur est associée à une probabilité plus élevée de conserver un emploi permanent (EDI).

5. Ajustement du modèle de Kaplan-Meier par genre et sorti du supérieur.

```
km_fit_Q1_SUPER <- survfit(Surv(duree, EDI) ~ Q1+SUPER, data = base)

# Résultats du modèle
summary(km_fit_Q1_SUPER)

# Visualisation de la fonction de survie estimée par zone géographique
ggsurvplot(km_fit_Q1_SUPER, data = base, xlab = "durée", ylab = "Probabilité
é de survie",
           conf.int = TRUE, risk.table = TRUE, pval = TRUE, ggtheme = theme
_minimal())
```



Interprétations : En supposant tous les autres facteurs constants, les femmes tous comme les hommes ayant atteint le niveau supérieur accédant plus rapidement à un EDI comparativement à ceux ou celles qui n'ont pas atteint le niveau supérieur. Par ailleurs les femmes du supérieur accédant plus vite que les hommes du supérieur.

```
# Test de Log-rank pour comparer Les courbes de genre et sorti du supérieur
survdif(Surv(duree, EDI) ~ Q1+SUPER, data = base)
```

```
## Call:
## survdiff(formula = Surv(duree, EDI) ~ Q1 + SUPER, data = base)
##
##
##                                     N Observed Expected (O-
E)^2/E
## Q1=femme, SUPER=Atteint niveau supérieur 21995      8580      6927
394
## Q1=femme, SUPER=Pas atteint niveau supérieur 11419      2917      4378
488
## Q1=homme, SUPER=Atteint niveau supérieur 17127      7044      5979
190
## Q1=homme, SUPER=Pas atteint niveau supérieur 19441      6047      7304
216
##                                     (O-E)^2/V
## Q1=femme, SUPER=Atteint niveau supérieur      591
## Q1=femme, SUPER=Pas atteint niveau supérieur      640
## Q1=homme, SUPER=Atteint niveau supérieur      271
## Q1=homme, SUPER=Pas atteint niveau supérieur      333
##
##  Chisq= 1394  on 3 degrees of freedom, p= <2e-16
```

Interprétations : Les résultats montrent des différences significatives dans la probabilité d'obtenir un EDI en fonction du sexe et du niveau d'éducation avec une p_value $p=2e-16 < 0.05$. Le sexe et l'obtention d'un niveau d'éducation supérieur ont une forte influence sur la probabilité d'obtenir un EDI, le niveau d'éducation supérieur augmentant généralement cette probabilité. De plus, les femmes et les hommes font face à des probabilités différentes.

X. REGRESSION DE COX

Afin d'éviter Le problème des valeurs manquantes nous allons définir nos modalités de référence pour chaque variable.

```
base$SUPER <- relevel(as.factor(base$SUPER), ref = "Atteint niveau supérieur")
base$phdip9 <- relevel(as.factor(base$phdip9), ref = "Non diplômé")
base$AP0 <- relevel(as.factor(base$AP0), ref = "contrat d'apprentissage non terminé")
base$AP2D <- relevel(as.factor(base$AP2D), ref = "ni travail ni stage")
base$ep15 <- relevel(as.factor(base$ep15), ref = "Par l'ANPE")
base$AP18 <- relevel(as.factor(base$AP18), ref = "pas de Contrat d'apprentissage antérieur")
base$GE01 <- relevel(as.factor(base$GE01), ref = "Uniquement dans votre région")
```

Ajustement du modèle de Cox

```
cox_model <- coxph(Surv(duree, EDI) ~ Q1 + phdip9 + SUPER + AP0 + AP2D + ep15 + AP6 + AP18 + GE01 + strateg, data = base)
```

Test de Schoenfeld pour vérifier Les risques proportionnels

```
cox.zph(cox_model)
```

```
##          chisq df      p
## Q1      3.15e-04  1 0.9858
## phdip9  8.65e+00  6 0.1942
## SUPER   4.07e+00  1 0.0436
```

```
## AP2D      4.10e+00  1 0.0429
## ep15      2.51e+01  8 0.0015
## AP6       6.46e-02  3 0.9957
## AP18      6.27e-01  1 0.4285
## GE01      8.20e+00  4 0.0845
## strateg  1.78e+01 12 0.1225
## GLOBAL    6.16e+01 37 0.0067
```

Les variables SUPER, AP2D ep15 montrent des violations des risques proportionnels ($p < 0.05$). Pour ces variables, il est donc nécessaire d'ajuster le modèle pour tenir compte de cette non-proportionnalité, par exemple en utilisant des modèles de Cox stratifiés ou en ajoutant des termes d'interaction dans le modèle. Dans la suite nous utilisons le modèle de de Cox stratifié.

#Modèle de Cox stratifié

```
model_cox_strat=coxph(Surv(duree, EDI) ~ Q1 + phdip9 + AP0 + AP18 + GE01
+ SUPER + AP2D + ep15 + strata(Strat) , data = base)
```

Test de Schoenfeld pour vérifier Les risques proportionnels

```
cox.zph(model_cox_strat)
```

```
##          chisq df      p
## Q1         0.849  1 0.3567
## phdip9     6.716  6 0.3479
## AP18        0.255  1 0.6132
## GE01        5.702  4 0.2226
## SUPER       1.293  1 0.2556
## AP2D        4.104  1 0.0428
## ep15       23.035  8 0.0033
## GLOBAL    37.880 22 0.0189
```

Les variables AP2D et ep15 montrent toujours des violations des risques proportionnels ($p < 0.05$). Il faut noter que nous ne pouvons pas utiliser la fonction *step* à cause du traitement des cases vides que nous avons traité au préalable. Nous allons donc enlever les variables qui montrent des violations des risques proportionnels pour espérer avoir un meilleur modèle.

#Modèle final

```
model_cox_final=coxph(Surv(duree, EDI) ~ Q1 + phdip9 + AP0 + AP18 + GE01
+ SUPER + strata(Strat) , data = base)
```

Test de Schoenfeld pour vérifier Les risques proportionnels

```
cox.zph(model_cox_final)
```

```
##          chisq df      p
## Q1         1.77  1 0.18
## phdip9     5.45  6 0.49
## AP18        1.69  1 0.19
## GE01        5.38  4 0.25
## SUPER       0.76  1 0.38
## GLOBAL    15.16 13 0.30
```

Les résultats indiquent que pour chaque variable individuelle ainsi que pour le modèle global, il n'y a pas de preuve significative de violation de l'hypothèse des risques proportionnels ($p > 0.05$ pour toutes les variables).


```
# Visualisation des résidus de Schoenfeld
```

```
ggcoxzph(cox.zph(model_cox_final))
```

```
## Warning: Removed 40 rows containing missing values or values outside the scale range
```

```
## (`geom_line()`).
```

```
## Warning: Removed 76 rows containing missing values or values outside the scale range
```

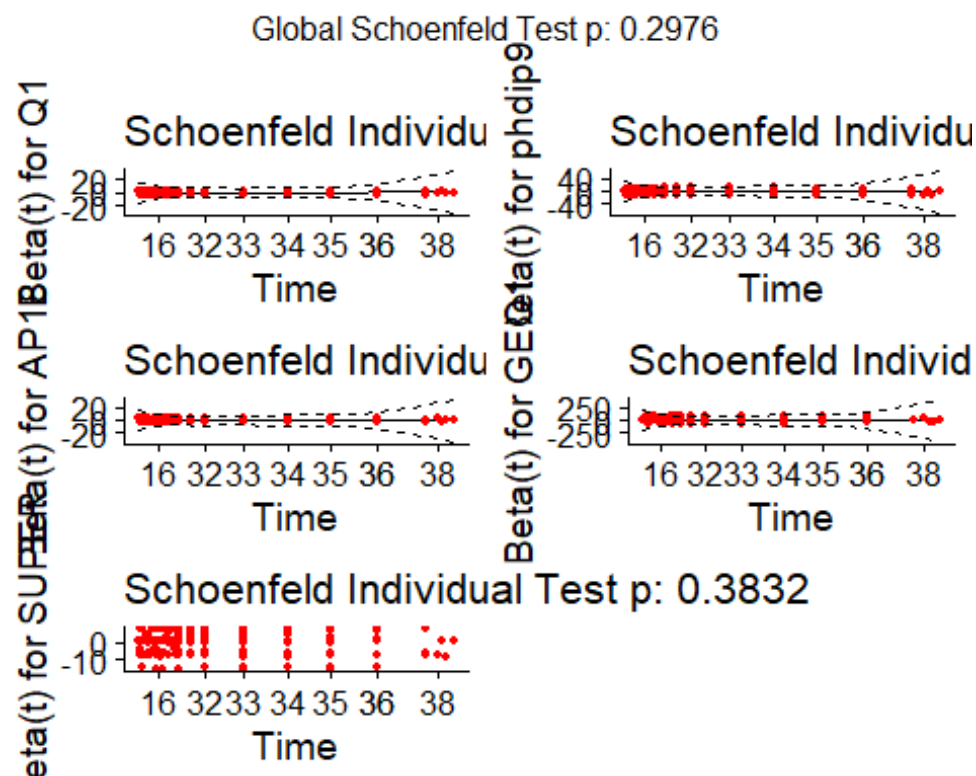
```
## (`geom_point()`).
```

```
## Warning: Removed 40 rows containing missing values or values outside the scale range
```

```
## (`geom_line()`).
```

```
## Removed 40 rows containing missing values or values outside the scale range
```

```
## (`geom_line()`).
```



Le test global de Schoenfeld montre que l'hypothèse de proportionnalité des risques est respectée ($p = 0.2976$). Les graphiques individuels des résidus de Schoenfeld pour chaque covariable confirment cette hypothèse, les p-values allant de 0 à 0.9983. Les points dans les graphiques sont répartis aléatoirement autour des lignes horizontales, indiquant qu'il n'y a pas de tendance systématique au fil du temps.

```
summary(model_cox_final)
```

```
## Call:
```

```
## coxph(formula = Surv(duree, EDI) ~ Q1 + phdip9 + AP0 + AP18 +
```

```
##     GEO1 + SUPER + strata(Strat), data = base)
```

```
##
```

```
## n= 875, number of events= 494
```

```
## (69107 observations effacées parce que manquantes)
```



```
##
##
##          coef exp(coef) se(coef)
z
## Q1homme          0.115538  1.122477  0.103270  1.1
19
## phdip9Bac          0.775371  2.171398  0.185093  4.1
89
## phdip9BTS-DUT-Santé-social  1.213926  3.366677  0.343358  3.5
35
## phdip9CAP-BEP-MC          0.445310  1.560974  0.162501  2.7
40
## phdip9DEA-DESS-Master-M2    1.672770  5.326904  0.358191  4.6
70
## phdip9Deug          NA          NA  0.000000
NA
## phdip9Doctorat          NA          NA  0.000000
NA
## phdip9Licence L3          1.345676  3.840782  0.379582  3.5
45
## phdip9Maîtrise M1, MST...    1.786525  5.968675  0.384597  4.6
45
## AP0contrat d'apprentissage terminé    NA          NA  0.000000
NA
## AP18Contrat d'apprentissage antérieur  0.222967  1.249780  0.107642  2.0
71
## GE01Dans plusieurs régions    0.085162  1.088893  0.149654  0.5
69
## GE01Dans toute la FRANCE    0.007145  1.007171  0.173979  0.0
41
## GE01En FRANCE ou à l'étranger -0.058225  0.943438  0.150751 -0.3
86
## GE01Uniquement dans une autre région  0.033943  1.034526  0.223651  0.1
52
## SUPERPas atteint niveau supérieur    0.106692  1.112591  0.283644  0.3
76
##          Pr(>|z|)
## Q1homme          0.263229
## phdip9Bac          2.80e-05 ***
## phdip9BTS-DUT-Santé-social  0.000407 ***
## phdip9CAP-BEP-MC          0.006137 **
## phdip9DEA-DESS-Master-M2    3.01e-06 ***
## phdip9Deug          NA
## phdip9Doctorat          NA
## phdip9Licence L3          0.000392 ***
## phdip9Maîtrise M1, MST...    3.40e-06 ***
## AP0contrat d'apprentissage terminé    NA
## AP18Contrat d'apprentissage antérieur  0.038324 *
## GE01Dans plusieurs régions    0.569318
## GE01Dans toute la FRANCE    0.967241
## GE01En FRANCE ou à l'étranger  0.699325
## GE01Uniquement dans une autre région  0.879370
## SUPERPas atteint niveau supérieur    0.706808
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
##                                     exp(coef) exp(-coef) lower .95 upp
er .95
## Q1homme                           1.1225      0.8909      0.9168
1.374
## phdip9Bac                         2.1714      0.4605      1.5107
3.121
## phdip9BTS-DUT-Santé-social       3.3667      0.2970      1.7177
6.599
## phdip9CAP-BEP-MC                  1.5610      0.6406      1.1352
2.146
## phdip9DEA-DESS-Master-M2         5.3269      0.1877      2.6399
10.749
## phdip9Deug                        NA           NA           NA
NA
## phdip9Doctorat                    NA           NA           NA
NA
## phdip9Licence L3                  3.8408      0.2604      1.8252
8.082
## phdip9Maîtrise M1, MST...         5.9687      0.1675      2.8087
12.684
## AP0contrat d'apprentissage terminé NA           NA           NA
NA
## AP18Contrat d'apprentissage antérieur 1.2498      0.8001      1.0121
1.543
## GE01Dans plusieurs régions        1.0889      0.9184      0.8121
1.460
## GE01Dans toute la FRANCE          1.0072      0.9929      0.7162
1.416
## GE01En FRANCE ou à l'étranger     0.9434      1.0600      0.7021
1.268
## GE01Uniquement dans une autre région 1.0345      0.9666      0.6674
1.604
## SUPERPas atteint niveau supérieur  1.1126      0.8988      0.6381
1.940
##
## Concordance= 0.622 (se = 0.016 )
## Likelihood ratio test= 75.77 on 13 df,  p=7e-11
## Wald test               = 80.98 on 13 df,  p=7e-12
## Score (logrank) test = 85.5 on 13 df,  p=1e-12
```

Les tests globaux du modèle, incluant le test de ratio de vraisemblance, le test de Wald et le test log-rank, sont tous hautement significatifs ($p < 0.001$), confirmant que les variables incluses dans le modèle contribuent de manière significative à l'explication de la survie.

Le modèle de Cox ajusté révèle que plusieurs niveaux d'éducation (phdip9) ont un impact significatif sur l'accès à un EDI. Les niveaux d'éducation Bac, BTS-DUT-Santé-social, CAP-BEP-MC, DEA-DESS-Master-M2, Licence L3 et Maîtrise M1 sont tous associés à une probabilité élevée d'avoir un EDI et des p-valeurs très significatives ($p < 0.01$), suggérant que des niveaux d'éducation plus élevés augmentent la chance d'avoir un EDI. En revanche, le sexe (Q1), la zone de recherche d'emploi (GE01) et le fait de ne pas avoir atteint le niveau supérieur (SUPER) ne montrent pas d'effet significatif sur la survie, leurs p-valeurs étant bien au-dessus du seuil de 0.05. Par ailleurs, la variable AP18 (Contrat

d'apprentissage antérieur) est également significative ($p = 0.038$), indiquant que les personnes ayant eu un contrat d'apprentissage antérieur ont une chance légèrement plus élevée d'avoir un EDI ($\exp(\text{coef}) = 1.25$).

Conclusion

Les résultats de cette étude ont permis de mettre en lumière les principaux facteurs influençant le délai d'accès au premier emploi durable pour les jeunes diplômés. Il ressort que le type et le niveau de formation, l'expérience professionnelle antérieure, et le lieu de résidence jouent un rôle significatif dans la rapidité d'accès à l'EDI. Ces informations sont cruciales pour les décideurs politiques et les organismes de formation, car elles permettent de mieux cibler les interventions visant à améliorer l'insertion professionnelle des jeunes. Alors, une approche intégrée prenant en compte ces différents facteurs peut potentiellement réduire le délai d'obtention du premier emploi durable, contribuant ainsi à une meilleure intégration des jeunes diplômés sur le marché du travail.