

RÉCIDIVISME DANS LES PRISONS DE L'ÉTAT DE L'IOWA AUX USA



LE JEU DE DONNEES

Analyse de survie sur les données de récidivisme dans l'Etat de l'Iowa, aux USA.

Durée du jeu de données : 5 années fiscales

Particularité des observations : les enregistrements de récidives s'étalent sur une durée de 3 ans

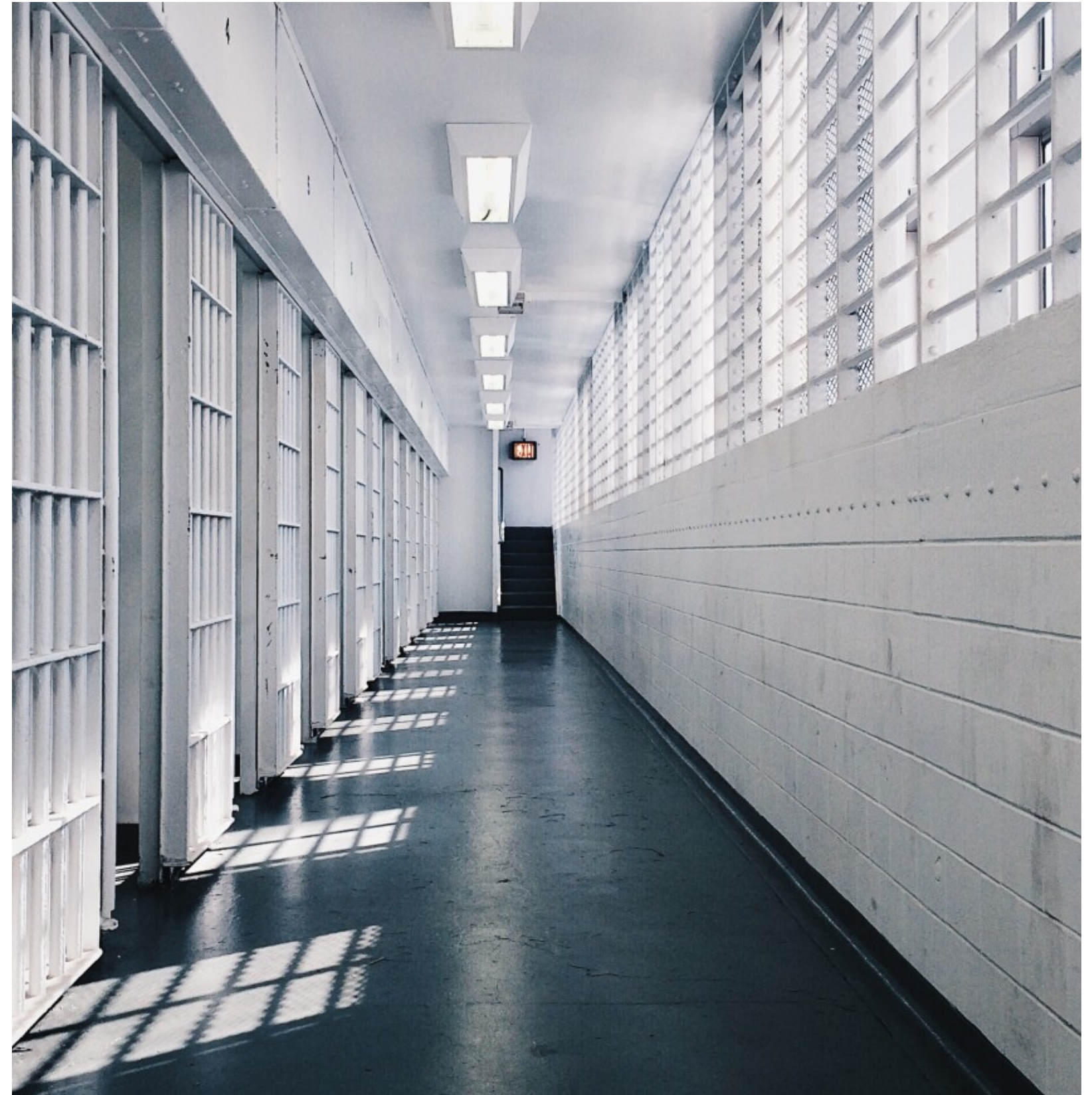
Année choisie : 2013

Taille du jeu : 3716 observations, 17 variables

Quels facteurs influent sur le récidivisme ?

L'analyse

- Préparation des données
- Estimation non paramétrique de Kaplan-Meier
- Modèle de Cox
- Vérification : critère d'Akaike



PRÉPARATION DES DONNÉES

```
#Charger les libraires  
library(readr)  
library(dplyr)  
library(survival)  
library(survminer)
```

- Chargement des libraires

```
#Charger les données  
recidive <- read_csv("C:/Users/Fujistone/Documents/Locole/Cours/statistiques  
2/Recidivism_stats_project/Recidive.csv")  
  
#Retirer les colonnes non pertinentes  
recidive = recidive[, -c(1,2,3,17)]
```

- Lecture du jeu de données
- Retrait des colonnes non pertinentes
- Renommer les colonnes(cf fichier .Rmd)

```
#Numériser la colonne "Return to Prison"  
recidive$`Return to Prison`[recidive$`Return to Prison` == "Yes"] = 1 #oui récidive  
recidive$`Return to Prison`[recidive$`Return to Prison` == "No"] = 0 #non pas récidive  
recidive$`Return to Prison` = as.numeric(recidive$`Return to Prison`)
```

- Numérisation de la variable de statut
-Transformer les variables en facteur(cf fichier .Rmd)

```
#Transformer les entrées de la colonne "Days to Return"  
recidive = mutate_if(recidive, is.numeric, ~replace(., is.na(.), 365*3))  
recidive$`Days to Return` = as.numeric(recidive$`Days to Return`)
```

- Retrait des NA de la variable temps

CENSURE

```
> library(asaur)
> table(recidive$return_prison)/nrow(recidive)
```

	0	1
	0.6967169	0.3032831

Censure à 69%
-> fort taux

30% de suivis sur les 3 ans

Estimation non paramétrique de Kaplan-Meier

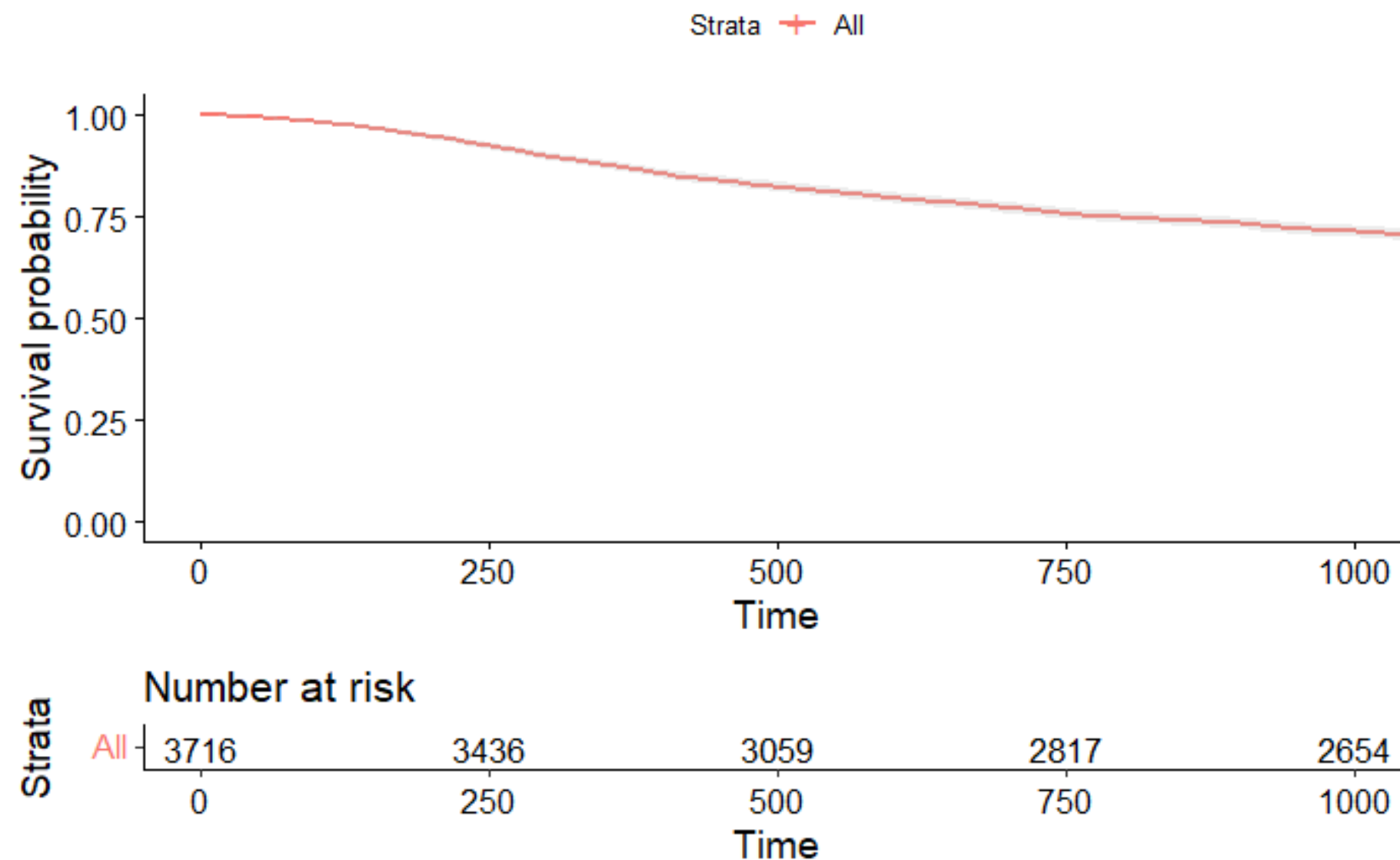
Fonction de survie globale
de Kaplan-Meier

- > Moins de 50% de prisonniers libérés ont récidivé au cours des 3 ans de suivi.
- > Risque décroissant constant
- > Récidive diminue en fonction du temps

```
library(survival)

KMO <- survfit(Surv(recidive$return_time, recidive$return_prison)~1,data = recidive)
print(KMO)
summary(KMO, times= seq(0,500,50))
```

```
#plot
library(survminer)
ggsurvplot(KMO,
            data = recidive,
            linetype = 1,
            conf.int = TRUE,
            risk.table = TRUE)
```



- > pas de diminution drastique de la survie au cours de la période de suivi (1095 jours)
- > Courbe décroissante quasi horizontale
- > Plus forte diminution aux alentours de 250 jours

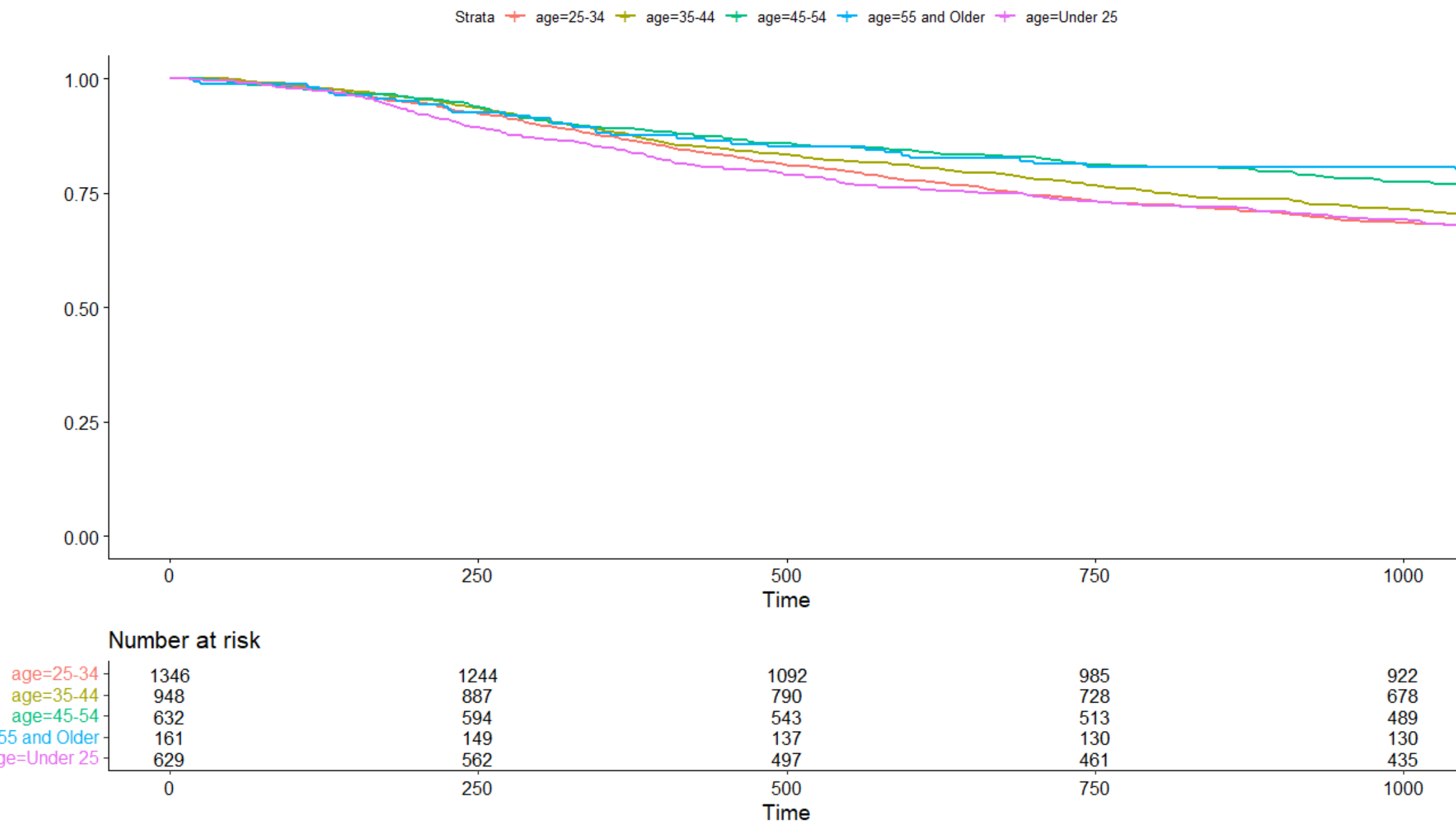
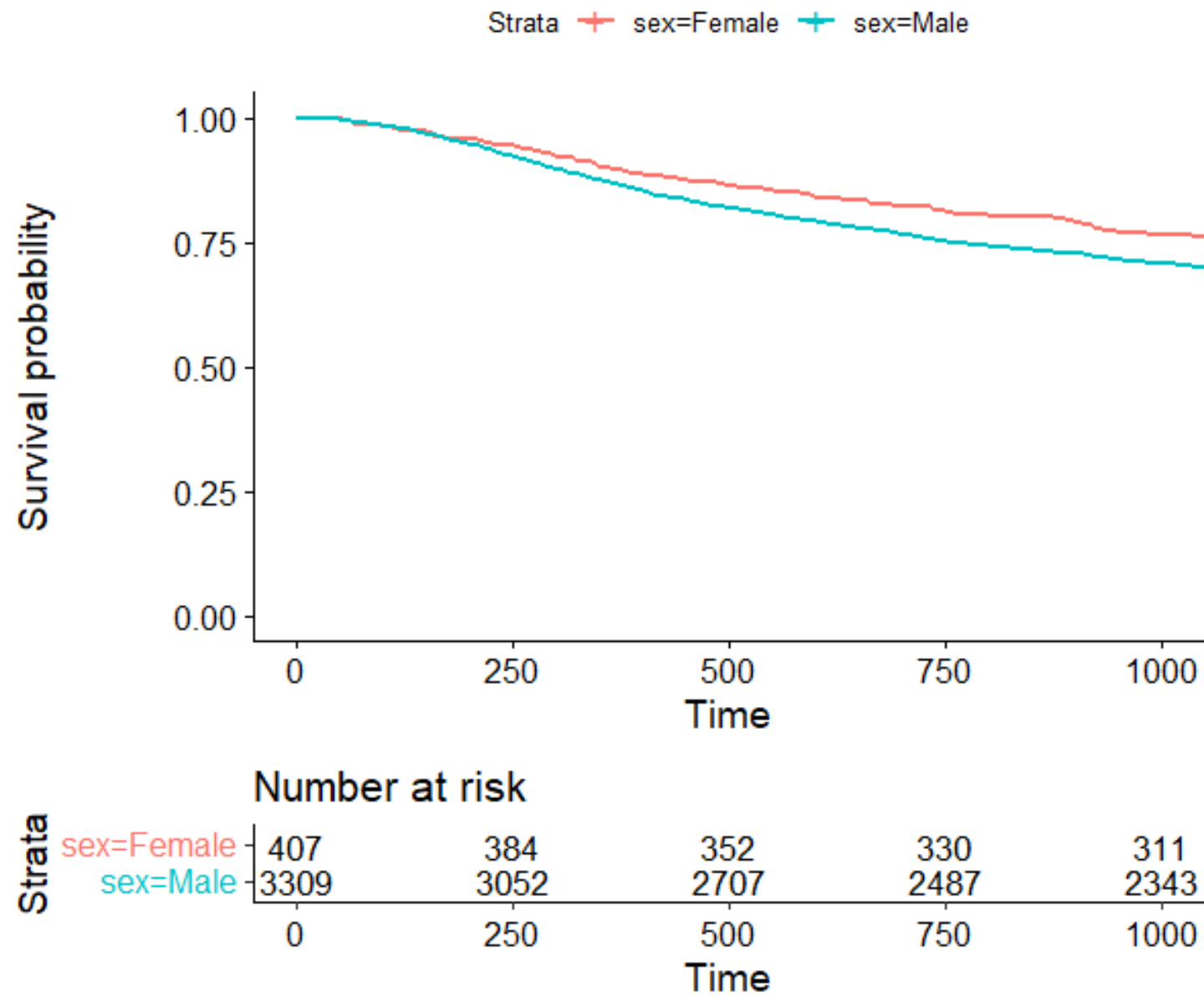
■ Estimation de Kaplan-Meier selon différentes variables

Estimation de Kaplan-Meier en fonction des autres variables du jeu données(cf fichier .Rmd)
-> la plupart ne sont pas significatives, on choisit de n'en garder que deux :

- l'âge au moment de libération
- le sexe

Selon l'âge

-> les moins de 55 ans sont plus susceptibles de récidiver
-> les moins 34 ans récidives le plus



Selon le sexe

-> les hommes récidivent plus que les femmes, mais l'écart est faible
-> l'écart se creuse aux environs de 200 jours

Selon le ciblage préalable des individus par le système carcéral

-> voir fichier .Rmd

TEST DU LOG-RANK

On pose l'hypothèse $H_0 = S_{\text{female}}(t) = S_{\text{male}}(t)$

```
survdif(formula = Surv(return_time, return_prison) ~ sex, data = recidive)
```

	N	Observed	Expected	(O-E) ² /E	(O-E) ² /V
sex=Female	407	102	128	5.431	6.13
sex=Male	3309	1025	999	0.698	6.13

chisq= 6.1 on 1 degrees of freedom, p= 0.01

- **p-value(sex) =**

0.01 < 0.05

- **Chi² = 6.1**

-> H₀ rejetée

- **p-value(age) =**

4e-05 < 0.05

- **Chi² = 25.6**

-> H₀ rejetée

```
survdif(formula = Surv(return_time, return_prison) ~ age, data = recidive)
```

	N	Observed	Expected	(O-E) ² /E	(O-E) ² /V
age=25-34	1346	443	402.5	4.07e+00	6.33e+00
age=35-44	948	290	289.9	7.54e-05	1.02e-04
age=45-54	632	152	199.1	1.11e+01	1.35e+01
age=55 and older	161	32	50.7	6.89e+00	7.21e+00
age=Under 25	629	210	184.8	3.43e+00	4.11e+00

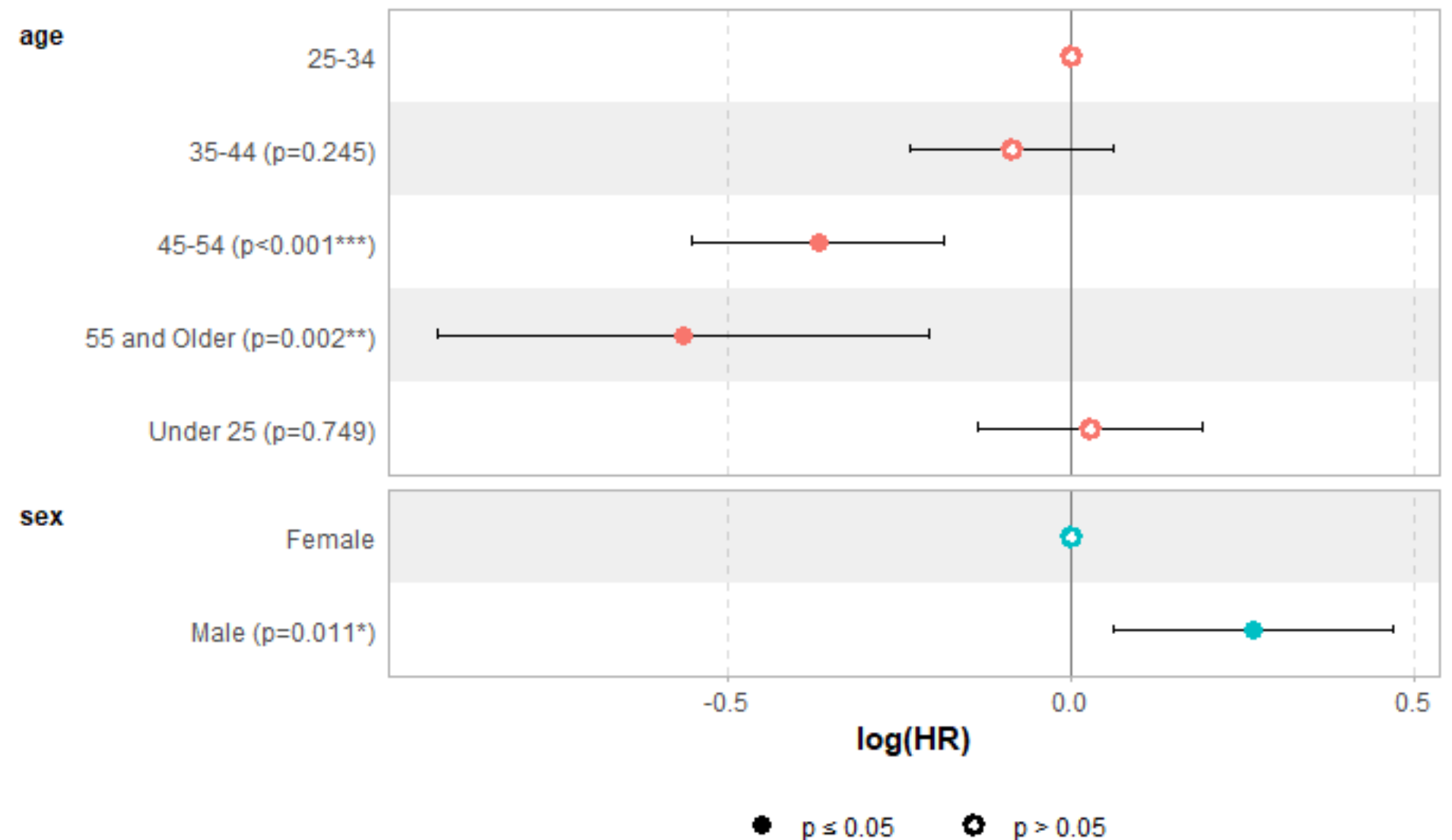
chisq= 25.6 on 4 degrees of freedom, p= 4e-05

REGRESSION DE COX

```
Variable      Units HazardRatio      CI.95      p-value
  age
    25-34      Ref
    35-44      0.92 [0.79;1.06]    0.24520
    45-54      0.69 [0.58;0.83]    < 0.001
    55 and older 0.57 [0.40;0.81]    0.00204
    Under 25    1.03 [0.87;1.21]    0.74945
  sex
    Female      Ref
    Male        1.30 [1.06;1.60]    0.01130
```

- L'âge(la jeunesse) est facteur de récidence
- Le sexe (masculin) est facteur de récidence

-> la figure confirme les résultats du tableau ci-dessus

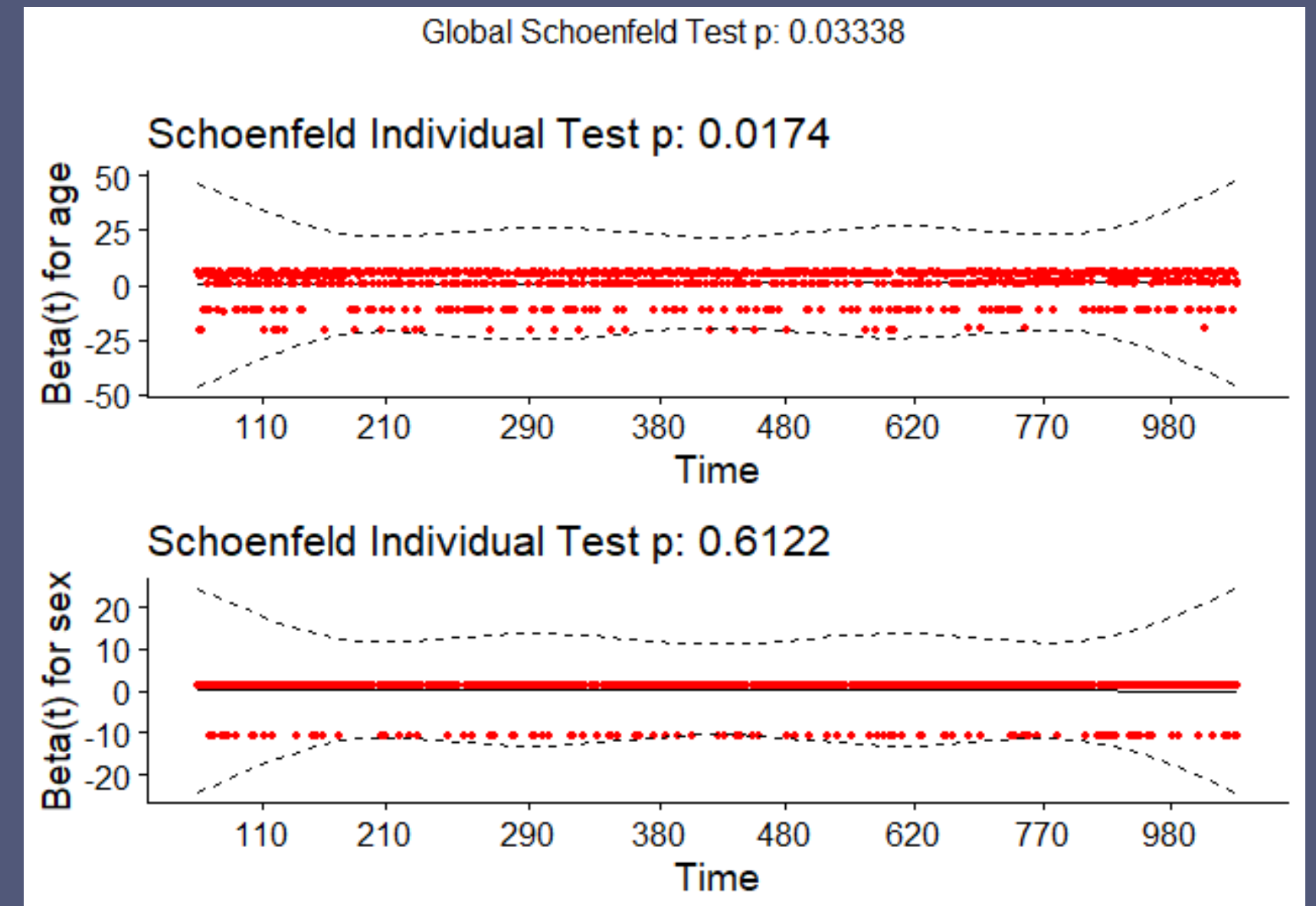


CRITÈRE D'AKAIKE ET VERIFICATION

Optimisation du modèle de
Cox grâce au critère
d'Akaike (Akaike
Information Criterion)

```
> cox1_recidive <- step(cox1)
Start:  AIC=18122.13
Surv(return_time, return_prison) ~ age + sex
```

	df	AIC
<none>		18122
- sex	1	18127
- age	4	18142



Résidus de Schoenfeld sont
constant en fonction du temps
-> variable sex et age sont
indépendantes

Conclusion

- Le sexe masculin est facteur de récidive
- Les jeunes (moins de 34 ans) sont également plus susceptibles de récidiver
- Le fait d'être ciblé au préalable par le système carcéral avant incarcération augmente de risque de récidive