

Dans le cadre d'un programme gouvernemental, on s'intéresse au nombre de semaines de paiements écoulé entre une perte d'emploi et un nouvel emploi pour des chômeurs et chômeuses âgé(e)s entre 18 et 65 ans, dans le but d'élaborer un programme de formation destiné au chômage de longue durée. La base de données chômage contient un échantillon de $n = 2000$ personnes inscrites à l'Assurance emploi entre janvier et novembre 2021.

- `duree` : nombre de semaines de prestations payées par le programme d'assurance emploi. Une valeur de `prestationmax+1` indique que la personne a épuisé ses prestations à `prestationmax` sans retrait du programme.
- `age` : variable catégorielle pour la tranche d'âge, soit [18,25), [25,55) et [55,65) ans.
- `formation` : variable catégorielle, soit aucune, secondaire, professionnel, collegial ou universitaire.
- `sexe` : variable indicatrice binaire; 1 pour femmes, 0 pour hommes.
- `retrait` : 1 si la personne bénéficiaire se désinscrit du programme avant la fin de la période maximale de prestation, par exemple si elle cesse de recherche un emploi ou si elle trouve un emploi, 0 sinon.
- `montant` : montant hebdomadaire des prestations, allant jusqu'à 55% de la rémunération hebdomadaire avant perte d'emploi jusqu'à concurrence de 547\$ par semaine.
- `nheures` : nombre d'heures cumulées à l'emploi avant d'être au chômage, soit un minimum de 14 semaines (420 heures) ou 22 semaines (700 heures) dépendamment du taux de chômage de la région et du nombre d'heures d'emploi assurables dans la période de référence.
- `prestationmax` : durée maximale pour les prestations, selon `pchomage` et `nsemaines`.

On s'intéresse au nombre de semaines qu'une personne passe au chômage. On suppose qu'aucun suivi n'est effectué par les employé(e)s de Service Canada au delà de la durée maximal de prestation.

1. (a) À quoi correspond la censure dans l'étude?
(b) Créez un indicateur statut indiquant la censure, avec 0 pour censure à droite et 1 sinon et rapportez le nombre d'observations censurées.
2. Expliquez dans vos mots l'impact de la censure à droite (non-informative) sur l'inférence : qu'arrive-t-il si vous ignorez cette dernière et modélisez directement le temps au chômage sans ajustement aucun (par exemple, en calculant les statistiques descriptives de durée)?
3. Supposons que l'on veuille comparer les courbes de survie selon le sexe et la formation en utilisant l'estimateur de Kaplan–Meier :
 - Testez l'égalité des courbes de survie à l'aide du test du log-rang. Écrivez les hypothèses nulle et alternative, la valeur de la statistique, la valeur- p et la conclusion du test.
 - Produisez un graphique des courbes de survie superposées pour chacun des sous-groupes avec intervalles de confiances ponctuels à 95%.
 - Rapportez l'estimé de la survie à 30 semaines pour une femme de 62 ans qui dispose d'une formation technique (collégial).
4. Le modèle de régression à risques proportionnels de Cox permettrait également de tester cette hypothèse. Quelles variables explicatives devriez-vous inclure pour ce faire dans le modèle et sous quelle forme?
5. Expliquez les avantages et inconvénients liés à l'utilisation du modèle de Cox par rapport à la spécification de courbes de survie différentes pour chaque sous-groupe.

6. Dans le cadre de l'étude, il serait opportun de considérer l'effet d'autres variables sur le chômage. Construisez un modèle de Cox à risque compétitifs avec sexe, age, formation, montant ~~et montantmax, une variable binaire indiquant si le revenu éligible excédait 547\$ par semaine.~~
- Interprétez l'effet des variables age et de la rémunération hebdomadaire sur le risque ~~de ne plus être au chômage. quitter le marché du travail (ne plus être à la recherche active d'emploi).~~
 - Toute chose étant égale par ailleurs, classez le temps médian avant la sortie du ~~chômage marché du travail~~ selon la formation, du plus long au plus court.
7. Les modèles ajustés ne se généralisent pas à la population complète : pourquoi? *Indication : quels sont les critères pour l'éligibilité aux paiements de l'assurance emploi?* Expliquez en quoi l'estimation précédente est faussée à cause de ces omissions.