

Ce travail est à réaliser en équipe (minimum deux, maximum quatre personnes).

Vous devez remettre les documents suivants :

- votre rapport au format PDF
- votre code **R** ou un fichier Rmarkdown

Utilisez la convention de nomenclature `d5_matricule.extension`, où `matricule` est le matricule de l'étudiant(e) qui soumet le rapport et `extension` est un de `pdf`, `R` ou `Rmd`.

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 `chomage` contient un échantillon de $n = 2\,500$ personnes inscrites à l'Assurance emploi entre mars 2020 et décembre 2021.

- `duree` : nombre de semaines de prestations payées par le programme d'assurance emploi.
- `dureemin` : nombre de semaines de prestations déjà écoulées au début de la collecte de données.
- `age` : variable catégorielle pour la tranche d'âge, soit [18,25), [25,50) et [50,65) ans.
- `formation` : variable catégorielle, soit aucune, secondaire, professionnel, collegial ou universitaire.
- `sexe` : variable catégorielle binaire; soit femme ou homme.
- `retrait` : 1 si la personne bénéficiaire se désinscrit du programme avant la fin de la période maximale de prestation parce qu'elle cesse de chercher un emploi, (2) 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.

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

1. (a) À quoi correspond la censure à droite dans l'étude?
(b) Rapportez le nombre d'observations censurées.
2. Supposons que l'on veuille comparer les courbes de survie selon le sexe et la formation (conjointement) en utilisant l'estimateur de Kaplan–Meier :
 - (a) 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.
 - (b) Produisez un graphique des courbes de survie superposées pour chacun des sous-groupes.
 - (c) Rapportez l'estimé de la survie à 30 semaines pour une femme de 62 ans qui dispose d'une formation technique (collégial) à l'aide de votre modèle.
3. 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?

4. 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.
5. Dans le cadre de l'étude, il serait aussi opportun de considérer le temps avant que la personne ne retrouve un **emploi**. Construisez un modèle de Cox à risques proportionnels avec `sexe`, `age`, `formation`, `montant` et `montantmax`, une variable binaire indiquant si le revenu éligible égale ou excède 547\$ par semaine.
 - (a) Interprétez l'effet de la rémunération hebdomadaire sur le risque de retrouver un emploi, incluant le cas où le ou la bénéficiaire reçoit le montant maximum.
 - (b) Toute chose étant égale par ailleurs, classez le temps médian avant le retour à l'emploi selon la formation, du plus long au plus court.
 - (c) À l'aide du test de score, vérifiez le postulat de risque proportionnels. Rapportez la valeur- p et les conclusions des tests.