

7.1 On considère le nombre de semaines pendant lequel des mères allaitent leur nouveaux nés. La base de donnée allaitement contient les variables suivantes :

- *duree* : durée de l'allaitement (en semaines)
- *censure* : variable indicatrice pour la censure, 1 si le temps correspond à la fin de la période d'allaitement, 0 si le temps est censuré à droite
- *race* : race de la mère, soit blanc (1), noir (2) ou autre (3)
- *pauvrete* : est-ce que le revenu de la famille est inférieur au seuil de pauvreté? soit oui (1) ou non (0)
- *fumeur* : est-ce que la mère fumait à la naissance de l'enfant? soit oui (1) ou non (0)
- *alcool* : est-ce que la mère consommait de l'alcool à la naissance de l'enfant? soit oui (1) ou non (0)
- *agemere* : âge de la mère à la naissance de l'enfant
- *annee* : année de la naissance de l'enfant
- *scolarite* : nombre d'années de scolarité de la mère
- *prenat3mois* : absence de soins prénataux : 0 si la mère a requis des soins prénataux dans les trois premiers mois de la grossesse, 1 sinon.

- (a) Estimez et tracez les courbes de survie pour les deux niveaux de la variable du statut de fumeur en utilisant l'estimateur de Kaplan–Meier. Commentez sur les fonctions de survie estimées.

Solution

La fonction de survie de Figure 1 descend jusqu'à zéro : dans les deux strates, la plus grande observation est observée. Les courbes diffèrent entre environ 25 et 50 semaines, sinon les intervalles de confiance se chevauchent.

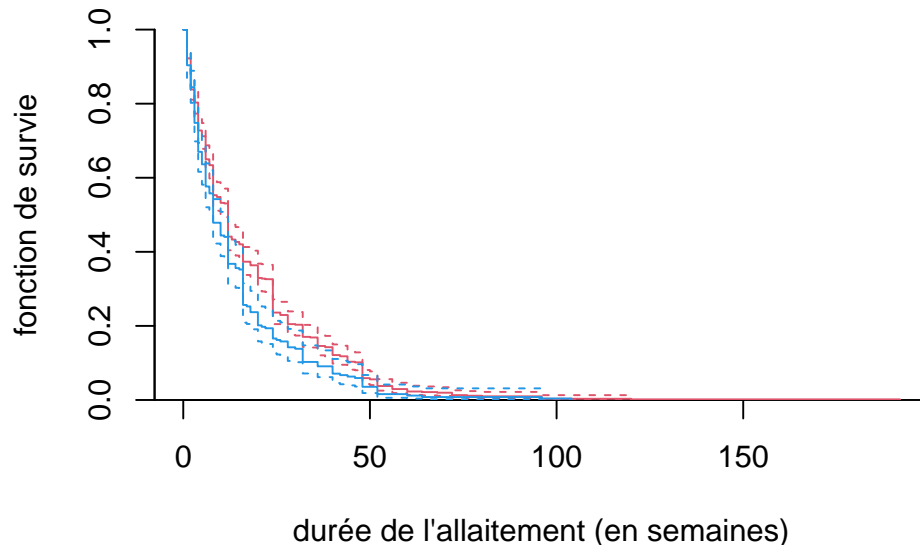


FIGURE 1 – Estimé de Kaplan–Meier de la fonction de survie pour les fumeuses (bleu) et les non-fumeuses (rouge)

- (b) Selon les résultats de la partie (a), quelle est la probabilité estimée qu'une mère non-fumeuse allaite pendant plus de 36 semaines? Qu'en est-il pour une mère fumeuse?

Solution

La probabilité de survie est de 0.1458 pour les non-fumeuses et de 0.0909 pour les fumeuses.

- (c) Quelle est la médiane et la moyenne du temps d'allaitement pour une mère non-fumeuse? Qu'en est-il pour une mère fumeuse?

Solution

Pour les non-fumeuses, la médiane et la moyenne sont 12 et 18.051 semaines, respectivement. Pour les fumeuses, la médiane et la moyenne sont 8 et 14.093 semaines.

- (d) Testez si les courbes de survie des mères non-fumeuses et fumeuses sont les mêmes.

Solution

La statistique du log rang vaut 10.09; en comparant au comportement d'une variable χ^2_1 , on rejette l'hypothèse nulle que les courbes de survie (ou la distribution de la durée d'allaitement) est la même pour les fumeuses et les non-fumeuses (valeur- p de 0.0015).

- (e) Ajustez un modèle de Cox à risques proportionnels pour évaluer l'impact du statut de pauvreté, du statut de fumeur, de l'âge de la mère et du nombre d'années de scolarité sur la fonction de risque. Écrivez la fonction de risque en fonction des estimés $\hat{\beta}$ et interpréter les estimés de ces paramètres.

Solution

La fonction de risque s'écrit

$$h(t; \mathbf{x}) = h_0(t) \exp(\beta_1 \text{pauvrete} + \beta_2 \text{fumeur} + \beta_3 \text{agemere} + \beta_4 \text{scolarite})$$

L'effet de l'âge de la mère et de la pauvreté n'impactent pas le risque, à niveau 5%.

- $\exp(\hat{\beta}_1) = 0.843$, donc les femmes qui vivent sous le seuil de la pauvreté ont un risque 15.7% moins élevé que les femmes plus aisées, *ceteris paribus*. Cela implique que les femmes pauvres ont probabilité plus élevée d'allaiter plus longtemps que les femmes plus aisées.
- $\exp(\hat{\beta}_2) = 1.218$; les fumeuses ont un risque 21.8% plus élevé que les non-fumeuses, toute chose étant égale par ailleurs. Le temps de survie est inférieur pour les fumeuses.
- Considérons deux femmes qui ont un an d'écart, *ceteris paribus*. Alors, le risque estimé est 1.019 fois plus élevé pour la femme plus âgée.
- Chaque année de scolarité additionnelle mène à une réduction du risque de 6%, toute chose étant égale par ailleurs; les femmes qui étudient davantage allaitent en moyenne plus longtemps.

- 7.2 Un commerce de chaussures de Montréal veut optimiser son inventaire afin de maximiser ses profits et fait appel à votre société de conseil. Les données chaussures contiennent les variables suivantes :

- statut : variable catégorielle, 0 s'il est vendu, 1 si l'article est toujours en stock, 2 s'il est déstocké.
- temps : temps de stockage de l'article (en mois).
- prix : prix de vente de l'article, arrondi à l'unité près.
- sexe : variable catégorielle, 0 pour modèle pour homme, 1 pour femme.

Notre objectif premier est d'estimer la fonction de survie du temps de l'article en stock.

- (a) Que représente la censure dans cet exemple?

Solution

Les options censurées correspondent aux paires de chaussure qui sont toujours en stock à la fin de la période d'observation ou qui sont invendues après 40 mois.

- (b) Estimez nonparamétriquement la fonction de survie du temps de stockage à l'aide d'un modèle de Kaplan-Meier et rapportez les estimés des quartiles.

Solution

Les quartiles sont 4, 7 et 11 mois.

- (c) Ajustez un modèle à risque proportionnel de Cox pour la durée de stockage en fonction du sexe et du prix de vente.

- i. Rapportez et interprétez les coefficients des variables. Est-ce que les effets estimés sont significatifs?

Solution

Les estimés (erreurs-type) des paramètres sont

- $\hat{\beta}_{\text{sexe}} = -0,164(0,026)$; le rapport de risque pour les femmes est $\exp(-0,164) = 0,848$ fois plus élevé que celui des hommes, pour un article au même prix. Les souliers pour femmes restent plus longtemps en stock.
- $\hat{\beta}_{\text{prix}} = -0,0136(9,59 \times 10^{-4})$; le rapport de risque de souliers pour hommes ou pour femmes est 0,986 fois celui d'une paire un dollar moins élevé, soit une diminution de 12,7% pour une différence de prix de 10\$.

Les deux coefficients sont significatifs selon les tests de Wald (valeurs- p inférieures à 10^{-4}). Les covariables sont aussi globalement significatives (statistique du rapport de vraisemblance pour $\mathcal{H}_0 : \beta_{\text{sexe}} = \beta_{\text{prix}} = 0$ de 248, valeur- p négligeable).