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 (les modèles invendus après 40 mois en magasins sont passés aux pertes et profits).
— `temps` : temps de stockage de l'article (en mois).
— `prix` : prix de vente réelle de l'article (avec rabais si applicable), arrondi à l'unité près.
— `sexe` : variable catégorielle, `0` pour modèle pour homme, `1` pour femme.

Notre objectif premier est d'estimer le temps qu'un article passe en stock avant d'être vendu.

7.1 Que représente la censure dans cet exemple ?

Solution
Les données censurées à droites 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.

7.2 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.

```
proc lifetest data=souliers method=km;
time temps*statut(1,2);
run;
```

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

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

Solution
Les estimés (erreurs-type) des paramètres sont

— $\widehat{\beta}_{\texttt{sexe}} = -0,164(0,026)$; 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.
— $\widehat{\beta}_{\texttt{prix}} = -0,0136(9,59 \times 10^{-4})$; 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 $\mathscr{H}_0 : \beta_{\texttt{sexe}} = \beta_{\texttt{prix}} = 0$ de 248, valeur-p négligeable).

(b) Tracez les estimés des courbes de survie d'une chaussure de l'année dont le prix de vente est 120$ pour les modèles pour homme et pour femme.

Solution
Voir la Figure 1 pour l'estimation de la survie basée sur le modèle à risque proportionnel de Cox ; la fonction de risque de base provient de l'estimateur de Kaplan–Meier.

On vous informe que, pour éliminer les invendus lors de l'arrivée de nouveaux modèles, l'entreprise offre une réduction de 20% après 15 mois.

7.4 Ajustez un modèle de Cox qui prendra en compte cette nouvelle information. Rapportez les estimés des paramètres de votre modèle ; est-ce que vos interprétations changent ?

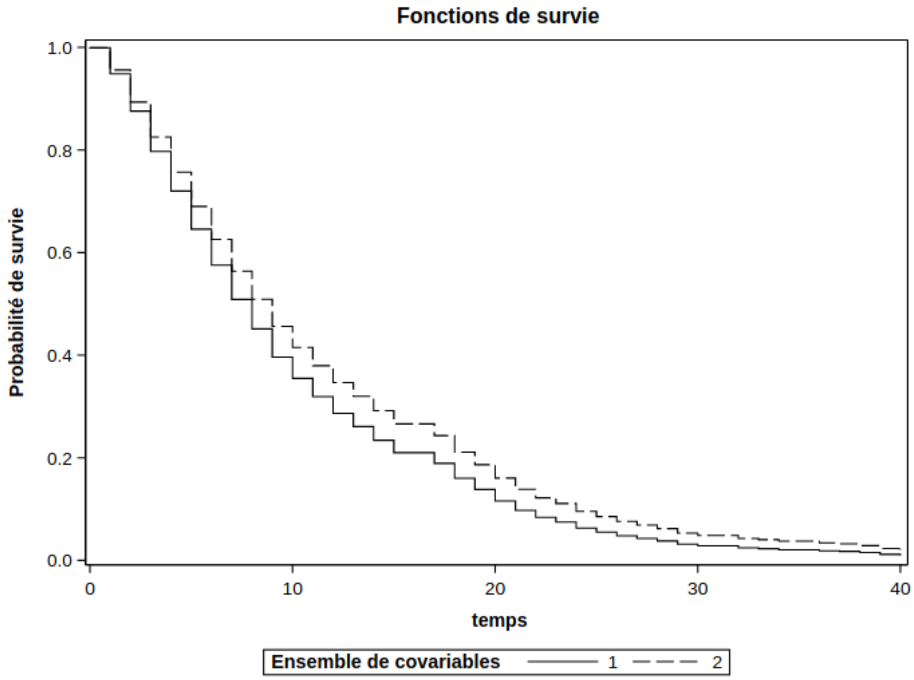

FIGURE 1 – Estimé de la fonction de survie pour une paire de chaussure de 120$ pour hommes (traitillés) et femme (complet)

Solution
Non ; les estimés sont toujours dans la marge d'erreur et leurs effets est toujours significativement non-nuls.

7.5 On pourrait considérer un modèle à risque non-proportionnels contenant une interaction entre le prix et le temps de manière à ajuster le même modèle. Expliquez comment cela pourrait être fait de manière à obtenir les mêmes estimés des paramètres.

Solution
On peut ajuster un modèle à risque non-proportionnel avec le terme d'interaction $\beta_{\texttt{prix}*t}\mathbf{1}_{t>15}\texttt{prix}$. En changeant les covariables, on obtient que $\beta_{\texttt{prix}*t} = -0{,}2\beta_{\texttt{prix}}$; cette contrainte peut être imposée en prenant comme modèle

$$h(t;\texttt{sexe},\texttt{prix}) = h_0(t)\exp\left\{\beta_{\texttt{sexe}}\texttt{sexe} + \beta_{\texttt{prix}}(\texttt{prix} - 0{,}2\texttt{prix}\mathbf{1}_{t>15})\right\}.$$