

# LE MODÈLE DE COX

Spécialisé dans le traitement statistique des études cliniques,  
le Département Biométrie de FOVEA vous propose  
au travers de ce fascicule quelques notions simples  
concernant le modèle de Cox.

*en  
pratique*

FOVEA

*Traitement statistique*

DES ETUDES CLINIQUES

# LE MODÈLE DE COX EN PRATIQUE

En épidémiologie étiologique, on est amené à **prendre en compte simultanément le rôle de plusieurs facteurs de risque dans la survenue d'une maladie**. En recherche clinique, on cherche à établir la relation liant une maladie à des facteurs pronostiques. **Les modèles multivariés permettent de représenter la variable étudiée en fonction de plusieurs autres variables**. Les principaux modèles utilisés sont la régression linéaire multiple, la régression logistique et le modèle de Cox. Le choix du type de modèle est lié à la nature des variables considérées.

- **La régression linéaire multiple**. La variable à expliquer est quantitative, sa distribution est normale.
- **La régression logistique**. La variable à expliquer est dichotomique, les variables explicatives peuvent être qualitatives ou quantitatives. Ce modèle permet de déterminer **la probabilité** de survenue de l'événement étudié en fonction des facteurs explicatifs.
- **Le modèle de Cox**. La variable à expliquer est dichotomique, les variables explicatives peuvent être qualitatives ou quantitatives. Ce modèle permet d'exprimer **le risque instantané** de survenue de l'événement en fonction des facteurs explicatifs.

## 1

### Principe généraux et définitions

Le modèle de Cox s'applique à toute situation où l'on étudie le **délai de survenue d'un événement**. Cet événement peut être la récurrence d'une maladie, la réponse à un traitement, le décès,... Les sujets entrent dans l'étude au fur et à mesure qu'elle se déroule. Pour chaque sujet, on connaît la date **du début de l'observation (date d'origine)**, la date **des dernières nouvelles** et **l'état** par rapport à l'événement étudié.

Les sujets pour lesquels on ne connaît pas l'état à la date de fin de l'étude ou ceux dont la réponse par rapport à la survenue de l'événement est négative constituent **des données censurées**.

**Les valeurs des variables explicatives**  $X_i$  sont notées pour chaque sujet à la date de son entrée dans l'étude. Ces variables peuvent être **qualitatives** ou **quantitatives**.

La variable considérée  $T$  est le **délai** entre la date d'**origine** et la date de **survenue de l'événement étudié**. Le modèle de Cox permet d'exprimer le **risque instantané** de survenue de l'événement en fonction de **l'instant  $t$**  et des **variables**

**explicatives**  $X_i$ . Ces variables peuvent représenter des facteurs de risque, des facteurs pronostiques, des traitements, des caractéristiques intrinsèques au sujet, ...

**$n$  Le risque instantané** de survenue de l'événement  $(t, X_1, X_2, \dots, X_p)$  représente **la probabilité d'apparition** de l'événement dans un intervalle de temps  $[t, t + \Delta t]$  sachant que l'événement **ne s'est pas réalisé** avant l'instant  $t$ .

Le modèle de Cox exprime  $(t, X_1, X_2, \dots, X_p)$  sous la forme :

$$h(t, X_1, X_2, \dots, X_p) = h_0(t) \exp \left\{ \sum_{j=1}^p \beta_j X_j \right\}$$

Cette formule appelle quelques commentaires.

- Le risque instantané se décompose en 2 termes dont l'un **dépend du temps  $t$**  et l'autre **des variables  $X_j$** .
- Si par exemple, les variables  $X_j$  représentent des **facteurs de risque** et si elles sont toutes égales à 0,  $\lambda(t)$  est le **risque instantané** de sujets ne présentant **aucun facteur de risque**.
- La forme de  $\lambda(t)$  n'étant pas précisée, c'est plutôt **l'association** entre les **variables  $X_j$**  et la **survenue** de l'événement considéré qui est l'intérêt central du modèle. Cela revient à **déterminer les coefficients  $\beta_j$** .
- Le rapport des **risques instantanés** de 2 individus dont les caractéristiques respectives sont  $(X_1, X_2, \dots, X_p)$  et  $(X'_1, X'_2, \dots, X'_p)$  est :

$$\frac{\lambda(t, X_1, X_2, \dots, X_p)}{\lambda(t, X'_1, X'_2, \dots, X'_p)} = \frac{\exp(\sum_j \beta_j X_j)}{\exp(\sum_j \beta_j X'_j)}$$

Ce rapport ne dépend pas du temps. De tels modèles sont dits à **risques proportionnels**. C'est une hypothèse importante du modèle de Cox.

## 2

### Interprétation des coefficients

**n Variable  $X$  dichotomique**.  $X$  prend les valeurs 0 ou 1 selon l'absence ou la présence de la caractéristique considérée. Le rapport des risques instantanés des sujets de la classe 1 par rapport à la classe 0 est :

$$\frac{\lambda(t, 1)}{\lambda(t, 0)} = e^{\beta_j}$$

Le coefficient  $\beta_j$  est donc le **logarithme du risque instantané** relatif de la **classe 1** par rapport à la **classe 0**.

**n De façon générale**, les coefficients  $\beta_j$  représentent **l'effet de la caractéristique  $X_j$  et la survenue** de l'événement. Si  $\beta_j$  est **nul**, la  $j$ -ième caractéristique n'a **pas d'influence** sur l'événement considéré. Si  $\beta_j$  est **positif** et si 2 sujets ne diffèrent que par la  $j$ -ième caractéristique, des **valeurs élevées** de la  $j$ -ième caractéristique sont associées à un **risque instantané plus élevé**. Inversement

si  $\beta_j$  est **négatif**, des valeurs élevées de la  $j$ -ième caractéristique sont associées à un **risque instantané plus faible**.

## 3

### Estimation et tests

Le principe pour le modèle de Cox est de **n'estimer que les coefficients  $\beta_j$** . On ne cherche pas à estimer  $\lambda(t)$ . Les estimateurs des  $\beta_j$  sont obtenus par **la méthode du maximum de vraisemblance**. Plus exactement, seule la partie de la vraisemblance comportant de l'information sur les coefficients  $\beta_j$  est retenue pour les calculs.

On parle de "**vraisemblance partielle**" ou de "**vraisemblance de Cox**".

On teste l'hypothèse  $H_0$  que le vecteur des effets  $(\beta_1, \beta_2, \dots, \beta_p)$  est nul. 3 tests peuvent être utilisés : le test du score, le test de Wald et le test du rapport de vraisemblance.

---

## LE MODÈLE DE COX EN PRATIQUE

---

# 4

### Conclusion

Chacun des modèles multivariés que nous avons évoqué en introduction est plus adapté à un contexte qu'à un autre. Le modèle de Cox est adapté aux données dont le délai de suivi est variable selon les sujets et aux données censurées. Si la

période de suivi est fixe et qu'il n'y a pas de données censurées, le modèle de régression logistique convient aussi bien que le modèle de Cox.

Dans la formalisation du modèle de Cox, nous avons présenté le modèle sans interaction entre les variables explicatives. Il est possible de tenir compte de **l'effet de l'interaction** dans l'écriture du modèle.

Le modèle de Cox repose sur des hypothèses précises.

Le risque instantané de survenue de l'événement considéré s'écrit comme le produit d'une fonction qui dépend du temps et d'une fonction qui dépend uniquement des caractéristiques du sujet. C'est l'hypothèse des risques proportionnels. C'est aussi un modèle multiplicatif. Le risque instantané de survenue de l'événement est multiplié par une constante quand on change la valeur d'une variable explicative.

---

