

# Méthodologie

## Définition de l'Effet Aléatoire $v_i$

L'effet aléatoire  $v_i$  capture l'effet d'hétérogénéité non mesurable qui influence les résultats dans chaque agglomération. Il peut être considéré comme une variation spécifique à l'agglomération qui n'est pas expliquée par les covariables observées dans le modèle (comme les facteurs socio-économiques).

## Estimation de l'Effet Aléatoire $v_i$ dans les modèles mixtes

Pour une agglomération donnée  $i$ , l'effet aléatoire  $v_i$  représente la composante non observée qui affecte le nombre attendu de cas d'un événement (par exemple, le cancer œsophagien) dans cette agglomération. Voici comment cet effet aléatoire est généralement défini et estimé.

### Structure de Variance

Dans le cadre d'un modèle mixte, l'effet  $v_i$  est généralement supposé suivre une distribution normale multivariée :

$$v_i \sim N(0, \sigma_v^2)$$

où  $\sigma_v^2$  est la variance des effets aléatoires. Cela signifie que la plupart des valeurs de  $v_i$  seront proches de zéro, certains pouvant être positifs (indiquant un risque plus élevé que prévu) et d'autres négatifs (indiquant un risque plus faible).

### Estimation de $v_i$

## Méthodes d'Estimation

Pour une agglomération donnée  $i$ , l'estimation de  $v_i$  peut être réalisée à l'aide des méthodes suivantes :

### Approche de Maximum de Vraisemblance (ML)

- Une fois que le modèle est ajusté, les effets aléatoires sont souvent estimés comme partie intégrante du processus d'optimisation de la vraisemblance. Cela se fait généralement en maximisant la vraisemblance marginale des données en tenant compte des effets aléatoires.
- Après avoir obtenu les estimations des paramètres du modèle, on peut dériver les estimations de  $v_i$  à partir des résidus du modèle.

### Approche Empirique

- Après l'ajustement du modèle, les valeurs de  $v_i$  peuvent être calculées à partir des résidus du modèle pour chaque agglomération. Cela implique de comparer les cas observés dans chaque agglomération à ceux prédits par le modèle, en tenant compte de l'effet des covariables.

## Utilisation des Effets Aléatoires

L'effet  $v_i$  est essentiel pour comprendre comment des facteurs non mesurés ou non observés influencent les résultats dans une agglomération  $i$ . Il permet aussi :

- D'ajuster pour la variabilité non expliquée par les facteurs observés.
- D'améliorer la précision des estimations de risque dans le modèle en tenant compte des corrélations spatiales.

En résumé, pour une agglomération donnée  $i$ ,  $v_i$  représente le composant aléatoire du modèle qui ajuste l'incidence observée à celle attendue, en tenant compte de la structure de dépendance spatiale et des effets non mesurés. L'estimation de  $v_i$  se fait généralement par des méthodes comme le maximum de vraisemblance ou l'analyse des résidus après ajustement du modèle.

## Effets Aléatoires Spatiaux Basés sur le Voisinage $w_i$

### Modèle de Voisinage

Dans le modèle avec effets aléatoires basés sur le voisinage,  $w_i$  représente la variation qui dépend des valeurs observées dans les zones voisines. Ceci est souvent formulé en utilisant un **modèle autoregressif conditionnel (CAR)**.

### Structure des Modèles CAR

Le modèle CAR peut être spécifié comme suit :

$$w_i \mid w_j, j \neq i \sim N \left( \frac{1}{n_i} \sum_{j \in \delta_i} w_j, \sigma^2 \right)$$

où :

- $n_i$  est le nombre de voisins pour l'agglomération  $i$ ,
- $\delta_i$  est l'ensemble des voisins de l'agglomération  $i$ ,
- $\sigma^2$  est la variance de l'effet aléatoire.

Ce modèle suppose que la valeur de  $w_i$  dépend en moyenne des valeurs observées dans les zones voisines, ce qui permet de modéliser la corrélation spatiale.

### Estimation des Paramètres

Pour estimer les paramètres de ces modèles (y compris  $v_i$  et  $w_i$ ), une approche courante consiste à utiliser des méthodes spécifiques dans le cadre de la **maximum de vraisemblance** ou des **approches bayésiennes**.

### Étapes d'Estimation

#### 1. Pseudo-vraisemblance

- On utilise la vraisemblance conditionnelle de  $Y$  étant donné les effets aléatoires, pour améliorer l'estimation des paramètres.
- Cette approche permet de simplifier l'estimation en contournant certaines intégrations complexes dans les modèles à effets aléatoires.

#### 2. Optimisation

- Les algorithmes d'optimisation, tels que **Newton-Raphson**, sont utilisés pour maximiser la vraisemblance.
- Pour des approches bayésiennes, des méthodes de simulation comme **MCMC (Markov Chain Monte Carlo)** sont couramment appliquées pour estimer  $v_i$  et  $w_i$ .

## Modèle de régression logistique

Elle modélise une réponse binaire ( $y \sim B(n, p)$ ), où  $p$  est la probabilité de succès :

$$P(y|n, p) = \binom{n}{y} p^y (1-p)^{n-y}$$

La probabilité  $p$  est reliée au prédicteur par la fonction logistique :

$$p = \frac{1}{1 + e^{-\eta}} \quad \text{où} \quad \eta = \beta_0 + \sum_{i=1}^m \beta_i x_i.$$

La log-vraisemblance est exprimée comme :

$$\ell(\beta) = \sum_{i=1}^n [y_i \log p_i + (1 - y_i) \log (1 - p_i)].$$

## Modèle de régression de Poisson

La régression de Poisson est un type de modèle linéaire généralisé (GLM) utilisé pour modéliser des données de comptage. La distribution de Poisson suppose que la variance est égale à la moyenne, ce qui est souvent approprié pour des données de comptage. Le modèle de base est :

$$\log(\mu_i) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip} = (X\beta)_i$$

où  $\mu_i$  est le nombre attendu de cas pour l'observation  $i$ ,  $X_{ij}$  sont les variables explicatives, et  $\beta_j$  sont les coefficients à estimer.

## Modèles de Régression de Poisson avec Effets Aléatoires

### Régression Poisson avec Effets Aléatoires Non Spatiaux

La formulation générale de ce modèle peut être écrite comme suit :

$$\log(\mu_i) = \log(E_i) + (X\beta)_i + u_i$$

où :

- $\mu_i$  est la moyenne du nombre de visites pour l'agglomération  $i$ ,
- $E_i$  est la taille de la population de l'agglomération  $i$  (terme d'ajustement de l'exposition),
- $X$  est la matrice de conception pour les facteurs socio-économiques,
- $\beta$  est le vecteur des coefficients à estimer,
- $u_i$  est l'effet aléatoire spécifique à l'agglomération  $i$ , supposé indépendant et suivant une distribution normale.

## Régression Poisson avec effets aléatoires spatiaux basés sur la Distance

Ce modèle prend en compte l'autocorrélation spatiale par l'introduction d'effets aléatoires spatiaux :

$$\log(\mu_i) = \log(E_i) + (X\beta)_i + u_i + v_i$$

où :

- $v_i$  est l'effet aléatoire spatial basé sur la distance, qui peut suivre un **processus gaussien** avec une certaine structure de covariance (définie plus haut).

## Régression Poisson avec effets aléatoires spatiaux basés sur le voisinage

Ce modèle est similaire au précédent, mais  $v_i$  est remplacé par un effet aléatoire basé sur le voisinage :

$$\log(\mu_i) = \log(E_i) + (X\beta)_i + u_i + w_i$$

où :

- $w_i$  représente l'effet aléatoire spatial basé sur le voisinage (par exemple, un **modèle autoregressif conditionnel (CAR)** qui capture l'influence des agglomérations voisines).

## Note sur l'Estimation

Pour tous ces modèles, l'estimation des paramètres est réalisée à l'aide de la **méthode de pseudo-vraisemblance**, qui prend en compte les structures autocorrélées.

De plus, pour la comparaison des modèles, on utilise les critères suivants :

- **AIC (Akaike Information Criterion)** : évalue la qualité du modèle en pénalisant les modèles trop complexes.
- **BIC (Bayesian Information Criterion)** : similaire à l'AIC, mais avec une pénalisation plus forte sur la complexité du modèle.

Ces critères permettent de choisir le modèle le plus adapté en fonction des données disponibles.