Statistique et aide à la décision
Session 1

David Causeur
Agrocampus Ouest
IRMAR CNRS UMR 6625

Plan

1. Modèle de risque

2. Comparaison inter-groupe d’un risque

Facteur de risque

Accueil > Sciences > Groupe sanguin et coronavirus, un hasard génétique

Groupe sanguin et coronavirus, un hasard génétique

par **Sophie Bécherel** publié le 21 mars 2020 à 8h06

D'après une étude chinoise, les personnes de groupe sanguin O sont mieux immunisées contre le coronavirus que les autres groupes. Elles ont un risque d'infection 33% moindre. A contrario, les personnes de groupe A ont 20% de risque supplémentaire d'être infectées. Cette inégalité s'explique par l'action des anticorps.

Facteur de risque

medRxiv preprint doi: https://doi.org/10.1101/2020.03.11.20031096.this version posted March 27, 2020.

Relationship between the ABO Blood Group and the COVID-19 Susceptibility

Jiao Zhao[1,#], Yan Yang[2,#], Hanping Huang[3,#], Dong Li[4,#], Dongfeng Gu[1], Xiangfeng Lu[5], Zheng Zhang[2], Lei Liu[2], Ting Liu[3], Yukun Liu[6], Yunjiao He[1], Bin Sun[1], Meilan Wei[1], Guangyu Yang[7,*], Xinghuan Wang[8,*], Li Zhang[3,*], Xiaoyang Zhou[4,*], Mingzhao Xing[1,*], Peng George Wang[1,*]

Facteur de risque

We collected and ABO-typed blood samples from 1,775 patients infected with SARS-CoV-2, including 206 dead cases, at the Jinyintan Hospital in Wuhan, Hubei province, China. Another 113 and 285 patients with COVID-19 were respectively recruited from Renmin Hospital of Wuhan University, Hubei province and Shenzhen Third People's Hospital, Guangdong province, China. The diagnosis of COVID-19 was confirmed by a positive real-time reverse transcriptase polymerase-chain-reaction test of SARS-CoV-2 on nasal and pharyngeal swab specimens from patients. Two recent surveys of ABO blood group distribution of 3,694 normal people from Wuhan City and 23,386 normal people from Shenzhen City were used as comparison controls for the Wuhan and Shenzhen patients with COVID-19, respectively[5-6]. Statistical analyses were performed using chi-squared test. Data from different hospitals were meta-analyzed using random effects models, with calculation of odds ratio (OR) and 95% confidence interval (*CI*). Statistical analyses were performed using SPSS

Facteur de risque

Ceci a-t'il un effet sur cela ?

Deux types de variables derrière cette question :

- cela est la **variable réponse** Y,
- ceci est la **variable explicative** X.

... les variations de Y sont-elles liées à celles de X ?

- `mod = lm(y~x,data=...)`

Facteur de risque

Le modèle linéaire répond à la problématique si :

- Y quantitative continue, par exemple un rendement
- X quantitative ou catégorielle.

Et si Y n'est pas une variable quantitative continue ?

Exemple : le statut, sain/malade, d'un patient

Plan

Effet groupe sur un risque

Soit $Y \in \{y_1, \ldots, y_K\}$ une variable réponse à K groupes

et $X \in \{x_1, \ldots, x_I\}$ une variable explicative à I groupes

Indépendance entre X et Y :

$$\mathbb{P}(Y = y_k \mid X = x_i) \quad = \quad \mathbb{P}(Y = y_k), \text{ pour tout } k, \text{ pour tout } i$$

Effet groupe sur un risque

Soit $Y \in \{y_1, \ldots, y_K\}$ une variable réponse à K groupes

et $X \in \{x_1, \ldots, x_I\}$ une variable explicative à I groupes

Indépendance entre X et Y :

$$\begin{aligned}
\mathbb{P}(Y = y_k \mid X = x_i) &= \mathbb{P}(Y = y_k), \text{ pour tout } k, \text{ pour tout } i \\
\frac{\mathbb{P}(Y = y_k, X = x_i)}{\mathbb{P}(X = x_i)} &= \mathbb{P}(Y = y_k), \text{ [Théorème de Bayes]}, \\
\mathbb{P}(Y = y_k, X = x_i) &= \mathbb{P}(Y = y_k)\mathbb{P}(X = x_i).
\end{aligned}$$

On peut tester ces égalités à partir de la table de contingence.

Test du χ^2 de Pearson

Soit $Y \in \{y_1, \ldots, y_K\}$ une variable réponse à K groupes

et $X \in \{x_1, \ldots, x_I\}$ une variable explicative à I groupes

Table de contingence de X et Y :

	y_1	$\ldots$	y_K	**Total**
x_1	n_{11}	$\ldots$	n_{1K}	$n_{1\bullet}$
x_2	n_{21}	$\ldots$	n_{2K}	$n_{2\bullet}$
$\vdots$	$\vdots$		$\vdots$	$\vdots$
x_I	n_{I1}	$\ldots$	n_{IK}	$n_{I\bullet}$
Total	$n_{\bullet 1}$	$\ldots$	$n_{\bullet K}$	n

▶ RStudio

Test du χ^2 de Pearson

Sous H_0 : X et Y sont indépendantes,

$$\mathbb{P}_{\mathrm{H}_0}(Y = y_k, X = x_i) \quad = \quad \mathbb{P}(Y = y_k)\mathbb{P}(X = x_i).$$

Soit $n^\star_{ik}$ le nombre attendu d'individu sous H_0 pour lesquels $X = x_i$ et $Y = y_k$:

$$\begin{aligned} n^\star_{ik} &= n\,\mathbb{P}_{\mathrm{H}_0}(Y = y_k, X = x_i), \\ &= n\,\mathbb{P}(Y = y_k)\mathbb{P}(X = x_i), \\ \hat{n}^\star_{ik} &= n\,\frac{n_{\bullet k}}{n}\,\frac{n_{i\bullet}}{n}, \\ &= \frac{n_{\bullet k}n_{i\bullet}}{n}. \end{aligned}$$

▶ RStudio

Test du χ^2 de Pearson

Statistique de test : distance du χ^2 de Pearson

$$D^2 = \sum_{i=1}^{I} \sum_{k=1}^{K} \left(\frac{n_{ij} - n^\star_{ij}}{\sqrt{n^\star_{ij}}} \right)^2$$

Si $n^\star_{ik} \geq 5$, alors approximativement, :

$$D^2 \sim \chi^2_{(I-1)(K-1)} \text{ sous } H_0.$$

Degrés de liberté : $(I-1)(K-1) = (IK-1) - ((I-1) + (K-1))$

- RStudio

Test du χ^2 de Pearson

Supposons que Y a deux modalités, $Y = \pm 1$

Evaluation du risque : odds d'un groupe

$$\text{odds}_i = \frac{\mathbb{P}(Y = +1 \mid X = x_i)}{\mathbb{P}(Y = -1 \mid X = x_i)}$$

Comparaison par groupe des niveaux de risque : odds-ratio

$$\text{odds-ratio}_{i|j} = \frac{\text{odds}_i}{\text{odds}_j}$$

▶ RStudio

Limite du test du χ^2 de Pearson

Test d'indépendance entre deux variables

or il est souvent nécessaire de prendre en compte des facteurs de confusion

Etude Covid19-groupe sanguin :

Hypothèse implicite : dans l'échantillon, les patients atteints de Covid19 ont la même exposition à tous les autres facteurs de risque, quelque soit leur groupe sanguin.