

Tests d'hypothèse et biostatistique

Saviez-vous que la **statistique** était la dernière venue des disciplines mathématique ? Cela dit, son champ d'action est extrêmement large. En effet, on la rencontre dans des disciplines aussi diverses que variées. Ici, nous nous intéresserons plus particulièrement aux tests d'hypothèse et à la biostatistique et nous verrons plusieurs de leurs applications qui combinent **biologie** et **statistique**.

- Biostatistique et statistique :

La **statistique** est l'ensemble des techniques d'interprétation mathématique appliquées à des phénomènes pour lesquels une étude exhaustive de tous les facteurs est impossible à cause de leur grand nombre ou de leur complexité. Le Petit Robert

La **biostatistique** est l'application de ces méthodes au domaine médical. Il peut s'agir du recueil, de l'analyse et du traitement des données recueillies lors d'études biologiques. **Grâce à la biostatistique, nous pouvons par exemple tester si**

- 1) un médicament est plus efficace qu'un autre.
- 2) une maladie est génétique ou non.
- 3) les cigarettes light sont moins cancérigènes que les standard,...



- Les tests d'hypothèse :

Un **test d'hypothèse** est une démarche qui vise à discriminer, à un certain niveau de confiance, entre deux hypothèses statistiques; l'hypothèse nulle H_0 et la contre-hypothèse H_1 .

But : Voir si la différence observée entre deux ou plusieurs groupes est due au hasard de l'échantillonnage ou si la différence est significative.

On peut soit rejeter l'hypothèse nulle soit ne pas la rejeter (sans pour autant l'accepter). Lorsqu'on rejette H_0 , on accepte « l'hypothèse alternative » désignée par H_1 , toujours à un certain niveau de probabilités.

Risque de première et deuxième espèce :

En procédant ainsi, nous nous exposons à 2 types de risques :

-Rejeter H_0 et accepter H_1 alors que H_0 est vraie (risque de première espèce ou risque alpha)

-Ne pas rejeter H_0 alors que H_1 est vraie (risque de deuxième espèce ou risque bêta)

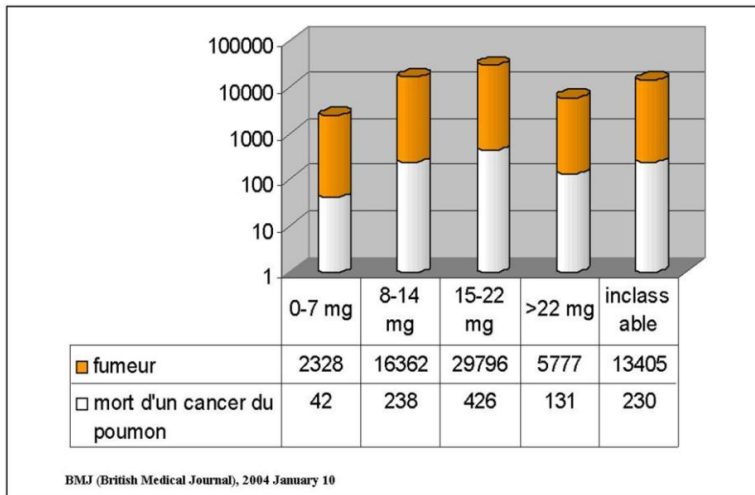
Il est évident qu'il n'existe pas de test sûr à 100%. On va néanmoins définir un certain niveau de confiance (le plus généralement à 95%) en dessous duquel on dira que le test est non significatif. Dans le cas contraire, la probabilité de commettre une erreur de première espèce est assez petite pour être considérée comme marginale.

- Démarche statistique en étapes:

- 1) Collecte de **données** qui, dans notre cas, remplissent les cases associées à chaque type de cigarette.
- 2) Réalisation d'un **test** qui, pour nous, sera un test d'homogénéité (en effet, nous voulons savoir si nos données sont homogènes, c'est-à-dire si les cigarettes light ne sont pas moins cancérigènes que les cigarettes standard). Pour cela, nous avons recours aux mathématiques et faisons quelques calculs.
- 3) Obtention de résultats chiffrés qui nous permettent de rejeter ou non H_0 et donc de **conclure**.

- Application :

Voici un graphique représentant le taux de cancer du poumon en fonction de la cigarette fumée.



- Calcul de test :

Pour notre problème (les cigarettes light sont-elles moins cancérigènes que les standard), nous allons tester si la sorte de cigarette fumée influence le risque de cancer de poumon et adoptons donc le modèle suivant :

H_0 : « les données sont homogènes ». } → Nous pratiquons donc un **test d'homogénéité**
 H_1 : « les données ne sont pas homogènes ».

dont la statistique de test est $T(X) = \sum_{k=1}^K \sum_{j=1}^J \frac{(n_{jk} - n_{.j} \hat{p}_k^0)^2}{n_{.j} \hat{p}_k^0}$ où $\hat{p}_k^0 = \frac{n_{.k}}{n}$ et qui, sous H_0 , se comporte comme une loi $\chi^2_{(K-1)(J-1)}$ où K est le nombre de colonnes et J le nombre de lignes.

A) Pour commencer, nous allons faire le test (au niveau 5%) pour toutes les catégories de cigarettes :
 $T = 25.94$ $\chi^2_{4;0.95} = 9.49$ (quantile de 95% pour une loi Chi-Carrée de paramètre 4).

La règle de calcul dit qu'on rejette H_0 si $T > \chi^2_{(K-1)(J-1);0.95}$, ainsi nous pouvons donc **rejeter l'hypothèse nulle H_0 d'homogénéité**.

En conclusion, il y a donc une grande probabilité pour que certaines cigarettes soient plus nocives que d'autres.

B) Néanmoins, il est assez évident que dans le tableau il y a un nombre qui se démarque clairement des autres, à savoir 17.49 dans la ligne >22 mg. Nous allons donc recommencer le test en supprimant le champ «>22 mg » :
 $T = 6.79$ et $\chi^2_{3;0.95} = 7.81$ (de paramètre 3).

Comme $T < 7.81$, nous ne pouvons **plus rejeter H_0** , ce qui signifie que, statistiquement, avec les données que nous possédons, il n'est pas prouvé que les résultats ne sont pas homogènes.

En conclusion, d'après ces deux tests, nous pouvons dire qu'**à part les cigarettes qui ont plus de 22 mg** (ce qui correspond aux cigarettes sans filtre), **il n'y a pas de cigarette moins nocive pour le cancer du poumon**.

- Avertissement

Gardez donc bien à l'esprit que les conclusions d'une enquête sans en avoir étudié le processus ne servent à rien. En effet, il est très facile de donner une interprétation erronée des résultats si on ne sait pas comment l'étude a été menée.

Et veillez également toujours à distinguer la signification statistique de la signification clinique. En effet, faire appel à l'une sans faire appel à l'autre peut s'avérer très risqué.