

Définition et expression graphique des paramètres biostatistiques

Didier Lepelletier¹, Philippe Vanhems²

1- Équipe opérationnelle d'hygiène, service de bactériologie-hygiène, CHU, Nantes
2- Département d'hygiène, épidémiologie et prévention, hôpital Edouard-Herriot, Lyon

Introduction

Le terme « statistique » est emprunté au latin moderne *statisticus* « relatif à l'État ». Il a d'abord désigné l'étude méthodique des faits sociaux définissant un État, par des procédés numériques (dénombrement, inventaires chiffrés, recensement, etc.). Dans sa théorie analytique des probabilités, Laplace, au début du XIX^e siècle, met en évidence le rôle que l'on peut en tirer pour l'étude des phénomènes naturels dont les causes sont trop complexes pour pouvoir être analysées individuellement.

La méthode statistique se présente alors comme une méthode de collecte, de présentation et d'analyse des observations relatives à des individus ou à des faits nombreux appartenant à un même ensemble dont on souhaite mettre en évidence certaines propriétés générales.

Le terme *biostatistique* est d'introduction récente. Il concerne l'utilisation de la méthode statistique dans le domaine biologique ou médical. La connaissance des paramètres biostatistiques est essentielle afin d'interpréter avec pertinence des résultats épidémiologiques. L'interprétation de ces résultats doit utiliser des méthodes simples, descriptives et graphiques. Le texte fondamental sur l'interprétation graphique des données a été publié par John Tukey en 1977. Cet auteur décrivait comment « utiliser les graphes, les papiers millimétrés et les couleurs afin d'analyser les données pour savoir ce qu'elles expriment... ». Cette approche est celle d'un détective à la recherche d'indices pour identifier des relations entre plusieurs variables. Dans le domaine de l'investigation des épidémies (cf. fiche méthodologique investigation d'une épidémie, Bulletin de la SFHH, décembre 2007), la construction graphique d'une courbe épidémique, d'un tableau synoptique et d'une cartographie du lieu incriminé permet également à l'épidémiologiste de suivre l'évolution, la durée, l'importance et la fin de l'épidémie.

L'objectif de cette fiche méthodologique est de faciliter la compréhension et la lecture des paramètres biostatistiques en précisant leur définition et en représentant graphiquement les relations entre les variables.

Définitions

Échantillonnage

En pratique, il n'est pas possible d'étudier les caractéristiques d'une population dans son ensemble. L'usage est de ne considérer qu'une partie des sujets appartenant à cette population. Ce sous-ensemble est appelé **échantillon**. En théorie, pour qu'un échantillon soit représentatif de la population dont il est extrait, on réalise un tirage au sort. Ainsi, il est possible d'estimer la valeur non mesurée de la moyenne ou de la variance d'une population à partir d'un échantillon. Il s'agira alors d'utiliser un résultat d'estimation valide à des fins de généralisation à une population plus large dite population cible. Faute de représentativité (si l'échantillon ne possède pas les mêmes caractéristiques que la population que l'on souhaite étudier), les résultats obtenus sur un échantillon ne peuvent être généralisés à la population étudiée.

Statistique

Une statistique permet de décrire certaines caractéristiques d'un échantillon de la population. De nombreux paramètres statistiques existent et varient selon le type de variables et selon leur utilisation. Certains paramètres décrivent des informations de type tendance centrale et d'autres sont des paramètres de dispersion.

On citera

- **la moyenne** (valeur unique que devraient avoir tous les individus d'une population ou d'un échantillon pour que leur total soit inchangé, *somme des valeurs des individus/nombre d'individus*),
- **la médiane** (valeur à laquelle 50 % des valeurs observées sont inférieures, 50 % supérieures),
- **l'écart type** (quantifiant la dispersion des données autour de la moyenne),
- **la variance** (mesure arbitraire aléatoire servant à caractériser la dispersion d'un échantillon ou d'une population, *égale au carré de l'écart type*).
- **le mode** (ou valeur dominante désigne la valeur la plus représentée d'une variable quelconque dans une population),

Les valeurs minimales et maximales d'un échantillon représentent également des para-

mètres de dispersion.

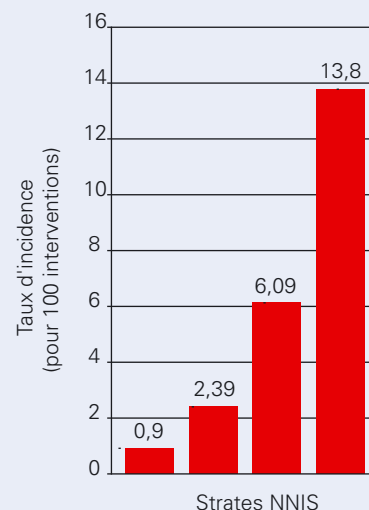
Du fait des fluctuations d'échantillonnage, l'estimation varie d'un échantillon à l'autre. Pour compléter l'estimation ponctuelle, on construit un intervalle de valeurs qui contiendra la valeur estimée avec un degré de confiance donné : c'est l'**intervalle de confiance à 95 %**. Si l'expérience est renouvelée 100 fois, dans 95 cas, la valeur se situera dans l'intervalle de confiance.

Type de variable étudiée et représentation graphique

Les variables qualitatives

Elles comprennent les variables nominales et ordinales. Les variables nominales ont deux modalités (homme/femme par exemple), ou plus de deux modalités et sont représentées généralement sous forme de « **camembert** » ou **secteur**. S'il y a une structure d'ordre entre les catégories d'une variable, on parle de variable ordinaire représentée habituellement sous forme d'**histogramme**. Dans la **figure 1**, l'histogramme montre la distribution du taux d'incidence des infections du site opératoire en fonction du score de risque infectieux NNIS. Les variables qualitatives sont généralement représentées sous forme de pourcentages ou de proportions.

Figure 1 - Exemple d'histogramme. Répartition du taux d'incidence d'infections du site opératoire en fonction du score NNIS (Source : données RAISIN 1999-2004).



Les variables quantitatives

Elles sont de deux types, soit **continues** lorsqu'elles peuvent prendre toutes les valeurs (âge, taille, poids, durée d'intervention), soient **discrètes** lorsqu'elles prennent des valeurs entières.

Dans le cas de variables quantitatives, le paramètre de tendance central le plus utilisé est *la moyenne*. Pour les variables ordinales, et dans certains cas pour les variables numériques, on utilise *la médiane*. On utilisera aussi la médiane en cas d'effectif réduit car ce paramètre est moins sensible aux valeurs extrêmes.

Figure 2 - Exemple de courbe. Expression de la consommation moyenne annuelle des produits hydro-alcooliques de 28 établissements de la région Ouest (Source : données C-CLIN Ouest).

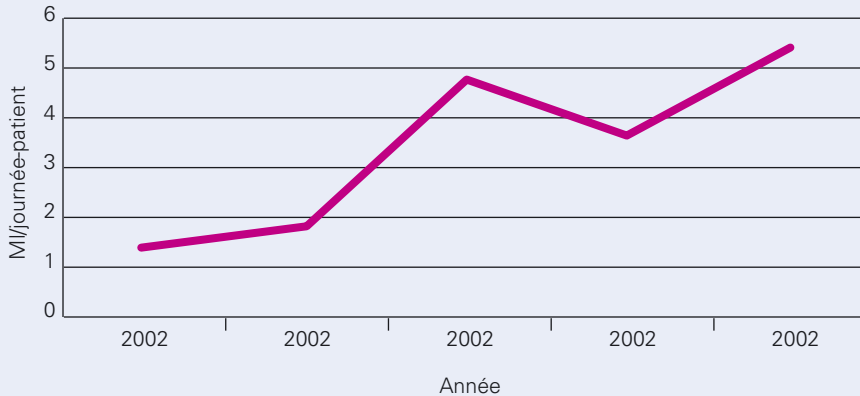


Figure 3 - Exemple d'histogramme. Expression de la durée d'intervention chirurgicale pour pontage aorto-coronarien (source : surveillance RAISIN, données CHU Nantes, avril-juin 2006).

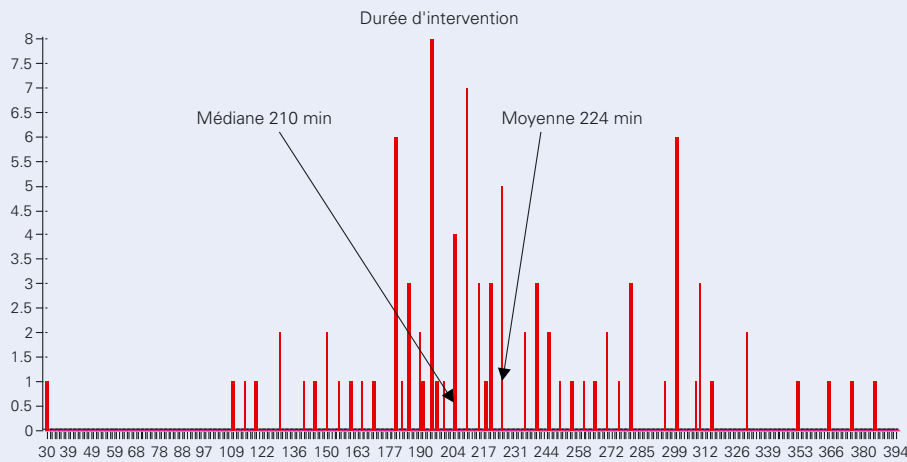
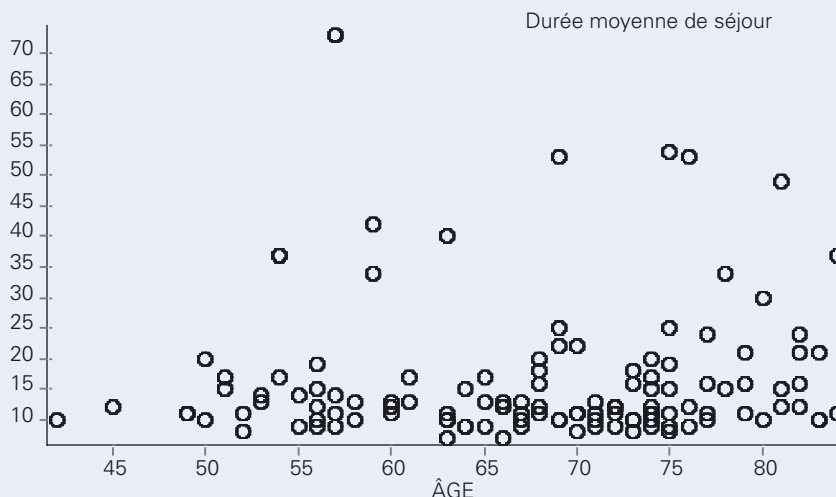


Figure 5 - Exemple de nuage de point. Expression de la durée moyenne de séjour pour pontage coronarien en fonction de l'âge (Source : surveillance RAISIN, données CHU Nantes, 2006).

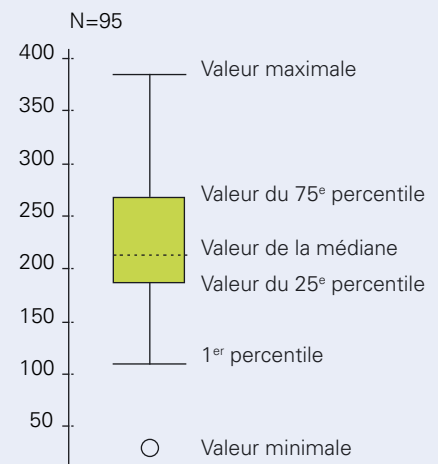


mes. Deux autres paramètres permettent de mesurer la dispersion des données, la distance interquartile et l'écart type représentés sous forme de courbe, d'histogramme ou de *box plot* (ou boîte à moustache).

Dans la **figure 2**, la courbe montre l'évolution de la consommation moyenne annuelle de produits hydro-alcooliques entre 2002 et 2006. Ce graphe permet de visualiser concrètement et rapidement l'augmentation de la consommation au cours du temps, sous condition que celle-ci soit continue.

Dans la **figure 3**, l'histogramme mon-

Figure 4 - Exemple de box plot (boîte à moustache) avec définition. Distribution de la durée d'intervention chirurgicale pour pontage coronarien (Source : surveillance RAISIN, données CHU Nantes, 2006).



tre la distribution d'une série de valeurs de durée d'intervention. La durée d'intervention comprend des valeurs extrêmement larges, avec de nombreux patients ayant une durée d'intervention plus longue que 224 minutes (moyenne). La moitié des patients ont une durée d'intervention inférieure à 210 minutes (médiane). Cette figure décrit les variables de manière plus efficace et plus précise que la traditionnelle moyenne \pm deux déviations standard (écart-type).

La boîte à moustache est une autre manière de visualiser les variations d'une distribution. La **figure 4** montre la boîte à moustache de la durée d'intervention exprimée dans le précédent histogramme. Comme détaillée dans la figure 3, la médiane et les tendances centrales apparaissent comme des lignes horizontales dans la « boîte ». Les 25^e et 75^e percentiles (ou 1^{er} et 3^e quartiles) mesurent les limites inférieure et supérieure de la « boîte ». Par définition, la distance entre ces bornes, « l'étendue interquartile », inclut la position centrale des données (50^e percentile), et reflète la distribution des données. À l'extrémité du graphe, des lignes traduisent les valeurs extrêmes des variables; un point excentré peut représenter une valeur individuelle. Si la distribution des valeurs des variables est normale à travers le *box plot*, la médiane et la moyenne sont égales. À travers la lecture de la boîte à moustache, il est possible de dire si la distribution est symétrique autour de la médiane et où se trouvent la plupart des valeurs. Le *box plot* est une visualisation pertinente pour examiner les valeurs extrêmes, qui peuvent représenter de véritables valeurs ou des valeurs aberrantes (erreurs de saisie).

L'utilisation d'un nuage de points permet d'analyser graphiquement deux variables en même temps. Dans la **figure 5**, le nuage de points exprime la durée moyenne de séjour en fonction de l'âge du patient. À première vue, les points montent vers la droite, les patients âgés restant hospitalisés plus longtemps. Mais l'interprétation de cette distribution est délicate, car les points sont dispersés. Certains patients âgés ont des durées de séjour courtes. Un patient a une valeur extrême en haut

du graphe. La construction d'une droite au sein de ce nuage ne permettrait pas davantage d'interpréter la distribution des points. Plusieurs approches développées par Tukey permettent d'utiliser ce type de graphe avec pertinence (courbe au sein du nuage de point indiquant la tendance des variables).

Conclusion

L'analyse statistique exploratrice des données n'est pas nouvelle, mais les logiciels informatiques la rendent plus attractive. Ces techniques d'analyse graphique ne sont pas compliquées, et la publication de boîte à moustache est de plus en plus fréquente dans la littérature. Toute analyse de données nécessite d'en connaître les limites, et le recours à un statisticien est parfois nécessaire. Par ailleurs, un épidémiologiste utilisant des programmes d'analyse statistique sophistiqués a besoin de visualiser la distribution de ses données afin

d'interpréter ses résultats ou de comprendre une relation inattendue. La combinaison des deux approches représente un outil intéressant pour l'épidémiologiste dans le domaine de la santé, et notamment l'épidémiologie des infections nosocomiales.

Références

- 1- TUKEY JW. Exploratory data Analysis. Reading, MA, Addison Wesley; 1977.
- 2- MILLS JL. Data torturing. N Engl J Med 1993; 329: 1196-1199.

- 3- WILLIAMSON DF, PARKER RA, KENDRICK JS. The box plot: a simple visual method to interpret data. Ann Intern Med 1989; 110: 916-921.
- 4- Shelly MA. Exploratory Data Analysis: Data visualization or torture? Infect Control Hosp Epidemiol 1996; 17: 605-612.
- 5- BOUYER J, HÉMON D, CORDIER S, DERRIENNIC F, STÜCKER I, STENDEL B, CLAVEL J. Epidémiologie - Principes et méthodes quantitatives, éd. Inserm, 1996.
- 6- RUMEAU-ROUQUETTE C, BLONDEL B, KAMINSKI M, BRÉART G. Epidémiologie : Méthodes et Pratiques. Médecine-Sciences, éd. Flammarion, Paris 1995.

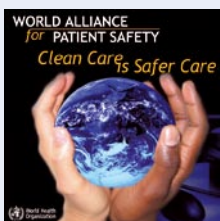
Journée nationale « Hygiène des mains » 23 mai 2008

Dans le cadre de son Alliance mondiale pour la sécurité du patient, l'Organisation mondiale de la santé (OMS) a lancé en 2005 le premier défi mondial pour la sécurité du patient avec un triple objectif : attirer l'attention sur les conséquences des infections nosocomiales, convaincre les pays de faire de la lutte contre ces infections une priorité et tester la mise en place des toutes récentes Recommandations pour l'hygiène des mains au cours des soins.

En France, dès juillet dernier, la cellule infection nosocomiale a mis en œuvre le projet d'organiser une journée nationale « Hygiène des Mains » en 2008, dans tous les établissements de santé. L'organisation de cette journée est pilotée par Lætitia May et Valérie Drouvot (Cellule infections nosocomiales - DHOS - DGS). Un comité scientifique et d'organisation multidisciplinaire a été constitué où la SFHH est représentée. Cette opération se déroulera à l'échelle des établissements de santé volontaires. Une charte signée par le directeur et le président de la CME engagera l'établissement dans la démarche. Un kit constitué d'outils pédagogiques sera mis à disposition par le ministère de la Santé.

Cette journée nationale aura lieu le vendredi 23 mai prochain. A cette date, Mme Roselyne Bachelot-Narquin, ministre de la santé, de la jeunesse et des sports signera publiquement la charte « un soin propre est un soin sûr » à l'occasion de la 60^e Assemblée mondiale de l'OMS.

MARTINE ERB



Adresse de l'image : http://www.who.int/gpsc/resources/Clean_CareLOGO2.jpg

Salle de congrès : Le Manège - Chambéry (Savoie)
Ce congrès est organisé sous l'égide du Pr René Moletta de l'Université de Savoie.

Introduction

Les rejets liquides (effluents) des établissements de santé ont des caractéristiques particulières par rapport aux effluents domestiques.

Ces établissements utilisent une grande variété de produits (antibiotiques, solvants, métaux lourds, radio-éléments... et les médicaments en général) mais aussi des produits d'hygiène et d'entretien, qui se retrouvent dans les eaux usées. Ces eaux peuvent être chargées aussi en micro-organismes (bactéries, champignons, virus... parfois résistants aux anti-infectieux). Tous ces produits se retrouvent au niveau de la station d'épuration qui le plus souvent met en œuvre des processus microbiologiques de dégradation. L'efficacité de ces procédés sur la dégradation et ou la rétention de ces produits est mal connue. Ces molécules se retrouvent parfois dans l'eau et dans les boues et, de fait, dans le milieu naturel.

Le devenir des produits de traitements hospitaliers lors de prises en charge ambulatoires relève de la même problématique.

L'Union Européenne a produit une directive assez restrictive sur les concentrations admissibles pour les rejets des stations d'épuration.

Le devenir des produits présents dans les rejets liquides des centres de soins (humain ou vétérinaire) est une question émergente qui doit être abordée de manière intégrée par les différents acteurs.

Objectifs

- Les objectifs de ce congrès sont :
- faire un état des lieux de la situation actuelle dans les différents domaines liés à la connaissance, à la gestion et au devenir de ces effluents.
 - faire se rencontrer les différentes parties prenantes de cette problématique :
 - les chercheurs,
 - les acteurs de la filière médicale (techniciens, médecins, pharmaciens, hospitaliers, infirmiers...),
 - les personnes responsables des aspects réglementaires et qui ont en charge de les faire appliquer,
 - les traiteurs d'eau,
 - les industries pharmaceutiques qui conçoivent et produisent les médicaments.

L'accent sera également mis sur la sensibilisation des étudiants, le congrès étant ouvert aux instituts de formation en soins infirmiers, à l'École des hautes études en santé publique, aux facultés de médecine et de pharmacie, et aux écoles d'ingénieurs sur l'environnement.

Pour en savoir plus, consulter l'annonce sur le site www.sfhh.net

DR JEAN CHARLES CÊTRE

13^e Forum international sur la qualité et la sécurité en soins de santé

Cette année, le 13^e Forum international sur la qualité et la sécurité en soins de santé aura lieu en France, du 22 au 25 avril, au Palais des Congrès de Paris.

Le Forum international rassemble les professionnels intéressés par l'amélioration de la qualité et la sécurité en soins de santé, avec une vocation très pédagogique et pratique.

La Société Française d'Hygiène Hospitalière a été sollicitée pour y organiser une

session qui traitera de la surveillance, du programme national de lutte contre les infections nosocomiales et des indicateurs, dans un esprit très didactique et interactif.

Les intervenants pour la SFHH seront les professeurs Jacques Fabry et Philippe Berthelot et le docteur Pierre Parneix. Les modérateurs seront les docteurs Olivia Keita-Perse et Serge Aho.

DR OLIVIA KEITA-PERSE