

Georges MION - Stéphane MÉRAT - Laurent GRASSER

Éléments indispensables de statistiques médicales : présentation des données

Le génie de la pensée précède souvent l'invention des outils d'investigation. C'est néanmoins le microscope qui a ouvert les portes de l'infiniment petit, le télescope celles de l'infiniment éloigné, et les inlassables calculs de l'ordinateur qui ont dévoilé l'infinie répétition des motifs du chaos. La méthodologie statistique est l'outil qui permet d'appréhender « l'incertain ».



Georges MION

Entre télescope et microscope, c'est là que nous sommes, en mer des tempêtes, au centre de l'écart, arc-boutés, cruels, opposants, hôtes indésirables.

René Char

Mots clés

Statistiques. Loi normale, box-plot. Échantillonnage. Randomisation.

Résumé

La finalité de l'utilisation des statistiques consiste à pouvoir dire si la variabilité d'un paramètre est le seul fruit du hasard. Les groupes étudiés doivent être représentatifs. Il est pour cela fait appel à la randomisation. Le nombre de sujets inclus doit aussi être suffisant, pour que l'étude soit valable.

Il est impensable de couvrir l'étendue de cette science en quelques pages. Aussi cette série d'articles aura-t-elle pour seule ambition de vous donner quelques définitions, des pistes pour savoir dans quelle direction chercher et surtout de vous mettre en garde vis à vis de la mauvaise utilisation de l'outil statistique, fréquemment dévoyé en recherche médicale^[1], bien qu'on observe une nette amélioration dans ce domaine^[2].

VARIABILITÉ, DISPERSION DES MESURES DE GRANDEURS PHYSIOLOGIQUES :

Tout paramètre physiologique, aussi régulé soit-il, admet des fluctuations. La finalité de l'utilisation des statistiques consiste à pouvoir dire (à l'issue d'un calcul) si la différence observée est le fruit du seul **hasard** (variabilité du paramètre étudié au sein d'une population donnée) ou non (échantillons issus de deux populations différentes). Dans certains cas, la réponse peut paraître évidente (mortalité des méningites bactériennes avant vs après la découverte des antibiotiques par exemple) ; la plupart du temps, la différence est suffisamment faible vis à vis de la variabilité spontanée du phénomène étudié pour que la simple observation ne permette pas de conclure avec certitude : les fluctuations dues au hasard introduisent un « **bruit de fond** », du « **flou** » dans notre vision médicale. Toute la finalité des tests statistiques sera finalement de comparer l'ordre de grandeur de la différence à évaluer, avec le niveau de ce bruit de fond.

ÉCHANTILLONNAGE :

Comme il est infaisable - ou trop coûteux - d'étudier une population toute entière, le chercheur (clinicien, biologiste, etc.) n'a à sa disposition que des groupes ou **échantillons** (de mesures, de patients) dont il espère qu'ils sont le reflet le plus fidèle de la (ou des) population(s) d'où ils ont été tirés (échantillonnage). Les tests d'hypothèses statistiques sont tous basés sur le postulat que les groupes étudiés sont bien **représentatifs** des populations correspondantes.

RANDOMISATION :

La **randomisation** (tirage au sort, à l'aide de tables de nombres au hasard, de dés ou d'une pièce de monnaie) est le seul garant de la **représentativité** d'un échantillon. En effet, la constitution de groupes historiques (avant/après), ou constitués d'après des critères arbitraires (initiale du nom, influencée par l'origine ethnique par exemple) ne garantit pas la **comparabilité** des groupes constitués. Exemple : on se pose la question de savoir si l'emploi de tel médicament amine vasopresseur améliore le pronostic de l'arrêt cardiaque. La différence apparue entre des groupes constitués *a posteriori* peut n'avoir rien à voir avec le traitement étudié : l'étiologie, le délai d'intervention, l'âge des patients peuvent expliquer les résultats. Pire, quand bien même la représentation de ces variables serait identique entre les deux groupes, des facteurs occultes peuvent être méconnus : compétence ou motivation des équipes d'intervention, biais introduits dans les soins au patient par la mise en route de l'étude, etc. Ainsi, seul le tirage au sort réalisé *a priori* peut répartir « **équitablement** » entre les groupes les fluctuations des

paramètres non étudiés mais susceptibles de « **polluer** » les résultats dans l'analyse finale (bruit de fond).

TYPES DE DONNÉES, DISTRIBUTION NORMALE :

Encore faut-il, pour que les groupes soient représentatifs d'une population, que le nombre de sujets inclus soit suffisant. Un tirage au sort sur une poignée de patients a peu de chance d'être représentatif de la population entière. En général, on considère que « **la loi des grands nombres** » s'applique aux données quantitatives à partir d'un effectif de 30 patients (Tableau 1 en page suivante).

Une étude publiée récemment dans *Science* montre précisément que c'est à partir de ce nombre que les criquets cessent de n'agir qu'en individus et commencent à adopter un comportement grégaire... Dans ces conditions, la répartition de paramètres physiologiques (âge, poids, taille, caractéristiques

« la randomisation est le seul garant de la représentativité d'un échantillon »

Types de données [3,4]	Caractéristiques	Exemples
Continues	Quantitatives, échelle de mesure à intervalles constants et zéro vrai	Débit cardiaque, glycémie
Continues sur échelle transformée	Quantitatives, échelle de mesure à intervalles constants, mais pas de zéro vrai	Température corporelle
Ordinales	Non mesurées, mais ordonnées ou rangées	Score ASA
Nominales	Qualitatives, classées en fonction d'un attribut	Survie ou décès, sexe

Tableau 1

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

métaboliques, etc.) suit en général une distribution en « cloche », dite loi normale, représentée par une courbe dite de Gauss (Fig. 1).

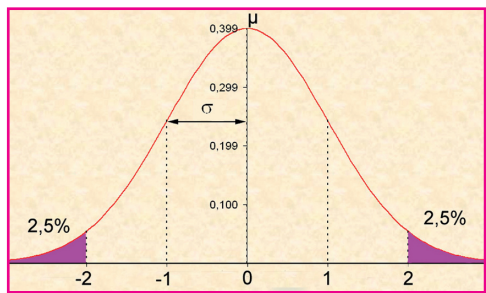
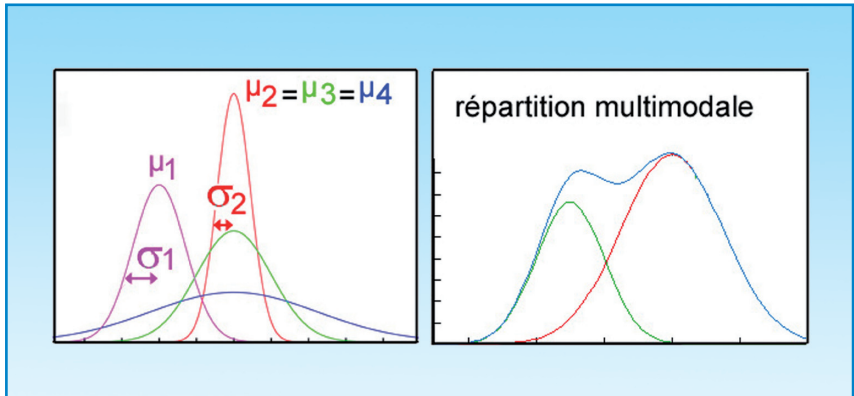


Figure 1 : Distribution normale (densité de probabilité), suivant une loi de Gauss-Laplace de moyenne μ et de variance σ^2 : La fonction est symétrique ; moyenne, médiane et mode sont confondus ; l'écart-type σ (ou déviation standard DS) mesure la distance entre l'axe de symétrie et les points d'inflexion. L'intervalle $\mu \pm \sigma$ contient 68% des valeurs, l'intervalle $\mu \pm 1,96 \sigma$ ($\approx \mu \pm 2 \sigma$) contient 95% des valeurs et l'intervalle $\mu \pm 3 \sigma$ plus de 99% des valeurs [3,5]. Ainsi, dans une population « normale », 5% seulement des individus (2,5% à droite et à gauche) s'écartent de plus de 2 DS de la moyenne.

PRÉSENTATION DES RÉSULTATS, SIGNIFICATION D'UNE MOYENNE, MÉDIANE, QUARTILES :

Deux populations de même moyenne ($\mu_2 = \mu_3 = \mu_4$) peuvent ne pas être identiques (Fig. 2). En effet, la moyenne « noie » les différences. Soit parce que la dispersion gaussienne des valeurs diffère autour de la même moyenne dans les populations étudiées (écarts-types ou variances différents), soit parce que des populations de même moyenne peuvent admettre des répartitions d'allure très différentes, non gaussiennes (plusieurs pics de fréquence par exemple). Ainsi, la première étape pour décrire une population ne devrait pas être le calcul des paramètres statistiques « de base » (moyenne, variance et

Figure 2.



écart-type), mais la vérification que le paramètre étudié suit bien une distribution normale (méthode graphique de Henry, test du Rankit ou utilisation de tests effectués par la plupart des logiciels dédiés). Dans ce cas, moyenne (arithmétique) et écart-type décrivent valablement la population, en en résumant les caractéristiques de centrage et de dispersion supposée symétrique, puisque la silhouette de la répartition est connue (courbe de Gauss). Dans ce seul cas s'appliquent les tests dits paramétriques, c'est à dire destinés à comparer des moyennes ou des variances.

Moyenne (théorique dans la population) :
 $\mu = \sum x_i / N$ (N = nombre de mesures)

Variance (théorique dans la population = moyenne des carrés des écarts) :
 $\sigma^2 = \sum (x_i - \mu)^2 / N = SCE / N$

La formule de la somme des carrés des écarts :
 $SCE = \sum x_i^2 - (\sum x_i)^2 / N$
 simplifie grandement les calculs [3].

Ecart-type, de même unité que la moyenne :
 $\sigma = \sqrt{\sigma^2} = \sqrt{\sum (x_i - \mu)^2 / N}$

Toutefois, dans un échantillon de n individus et de moyenne observée m, la variance s2, supérieure à σ^2 , doit être calculée par la formule :
 $s^2 = \sum (x_i - m)^2 / n - 1$
 et l'écart-type ou déviation standard

$DS = s = \sqrt{\sum (x_i - m)^2 / n - 1}$

(n-1) représente le nombre de degrés de libertés (ddl), c'est à dire le nombre de valeurs indépendantes utilisées pour calculer la variance de l'échantillon [3,5].

Tout écart e entre une valeur observée xi et la moyenne μ peut être exprimé avec l'unité σ : l'écart-réduit $\varepsilon = e/\sigma$ suit une loi normale particulière définie par $\mu = 0$ et $\sigma = 1$ (loi centrée-réduite). La table de l'écart-réduit donne la probabilité p pour que x sorte d'un intervalle donné centré sur la moyenne μ .

Pour une variable de distribution normale, on exprime ainsi un résultat sous forme d'une moyenne assortie de son écart type, de même unité (ce qui n'est pas le cas de la variance) :

Poids = 86 ± 7 kg

Toutefois, la dispersion exprimée par un écart-type de 7 kg est moindre pour un poids de 86 kg que pour un poids de 10 kg ; c'est la raison d'être du coefficient de variation exprimé sous la forme d'un pourcentage :

coef. Variation = σ / μ

Dès cette étape, certaines erreurs peuvent être évitées :

Erreur 1 :

poids = 86 ± 9,2 kg
 μ et σ doivent impérativement être exprimés avec le même nombre de décimales.

Erreur 2 :

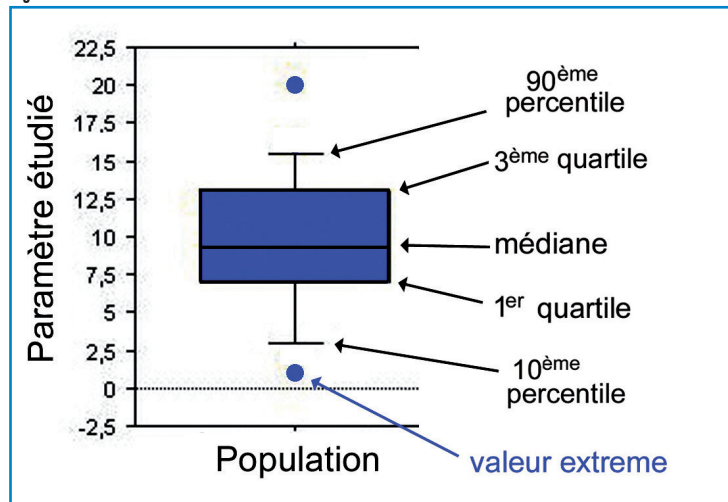
poids = 86,257 ± 9,293 Kg
 les 3 décimales calculées n'ont pas de pertinence compte tenu de l'imprécision de la mesure : cette forme de présentation sous-entend implicitement que votre pèse-personne est précis au gramme près. Souvent, il est judicieux d'arrondir, sous peine de non-sens.

Ainsi, pour une fréquence cardiaque, on préférera FC = 75 ± 12 à 75,3 ± 12,5 batt./min. Même remarque pour la mesure de la pression artérielle, dont la précision ne va pas au delà du mmHg^[4,5]. Les décimales doivent être évitées lors de l'expression de variables discontinues (ordinales), de même, pour un pourcentage, les décimales ne sont pertinentes que pour N ≥ 100.

Lorsque la distribution de la variable n'est pas normale, une description graphique des résultats (histogramme des fréquences) est plus judicieuse, et l'utilisation de tests dits non paramétriques, impérative (nous le verrons ultérieurement).

La moyenne comme indice de centrage devrait être alors remplacée par la médiane, « mesure du milieu »^[3]. Celle-ci, valeur qui sépare les n valeurs observées en deux groupes d'effectif égal, est valide quelle que soit la distribution observée (en cas de distribution gaussienne, elle se confond avec la moyenne). De même, lorsque la distribution n'est pas symétrique, la déviation standard n'est plus un indice de dispersion informatif et devrait être remplacée par l'utilisation des quartiles, valeurs au dessous desquelles se classent un quart (premier), la moitié (deuxième) et les trois quarts (troisième quartile) des valeurs observées. L'interquartile (écart entre premier et troisième quartile) est un indice simple et intéressant, bien que peu usité^[4]. On obtient la classique représentation en (box-plot, Fig 3).

Figure 3.



VALEURS EXTRÊMES :

Les valeurs extrêmes observées dans un échantillon (représentées par les points extrêmes sur le diagramme en box-plot) peuvent être mentionnées soit lorsqu'elles revêtent une signification physiologique (exemple, valeur d'une PIC de 25 mmHg) soit comme indices de dispersion d'une variable autour d'une médiane. On peut noter : le score de Glasgow médian était 8 (6-9) ou 8 (extrêmes : 6-9)^[4].

TRANSFORMATION DES DONNÉES :

Certaines distributions peuvent être rapprochées de la loi normale en utilisant des processus algébriques simples, comme l'expression sous forme de logarithme (distribution log-normale) ou de racine carrée ; en général, la distribution globale des données oriente les habitués vers tel ou tel type de transformation^[6]. Le recours aux tests non paramétriques est une forme particulière de transformation des données, où chaque valeur se voit substituée par son rang.

La moyenne géométrique est la racine nième du produit de chacune des mesures. Elle est plus adaptée pour décrire la tendance centrale de certains échantillons dont la distribution n'est pas gaussienne. C'est le cas de la concentration en ions hydrogène. En revanche, la moyenne arithmétique est mieux adaptée à l'expression de la tendance centrale du pH^[3].

ERREUR STANDARD SUR LA MOYENNE (ESM) OU DÉVIATION STANDARD ?

L'ESM ou erreur standard (ES) caractérise l'intervalle de dispersion possible de la moyenne théorique μ (dans la population d'où est extrait l'échantillon observé) par rapport à la moyenne m observée sur l'échantillon. Lorsque n tend vers l'infini, ESM tend vers 0, alors que DS tend vers σ ^[4].

$$ESM = DS / \sqrt{n}$$

Comme on peut surtout s'en rendre compte, si ESM et DS, qui sont souvent acceptées indifféremment par les revues scientifiques, peuvent se déduire l'une de l'autre (à condition que l'auteur fournisse la valeur de n, bien entendu), ESM permet de réduire la taille des « barres d'erreur » sur les graphiques

d'un facteur \sqrt{n} .

Dans ce but, l'utilisation d'ESM peut servir plus ou moins consciemment à affaiblir le sens critique des lecteurs en diminuant la dispersion apparente des résultats. Dans un souci d'éthique et de rigueur, il faut donc clairement mentionner n, le nombre de mesures effectuées et ne pas utiliser ESM pour de trop petits échantillons (en pratique pour n < 10). Enfin, pour diminuer ESM d'un facteur deux, il faut à l'évidence un échantillon quatre fois plus important.

Bibliographie

- Goodman NW, Powell CG. - *Could do better : statistics in anaesthesia research.* - Br J Anaesth 1998 ; 80 : 712 - 714.
- Horton NJ, Switzer SS. - *Statistical methods in the Journal.* - N Engl J Med 2005 ; 353 : 1977 - 1979.
- Fisher DM. - *Statistiques en anesthésie.* Pages 755-788 in : RD Miller (Ed). Anesthésie. Traduction dirigée par PJ Zetlaoui. Paris, Editions Flammarion Médecine-Sciences, 1996.
- Riou B, Pinaud M. - *Expression numérique des résultats.* - Ann Fr Anest Réan 1996 ; 15 : 49-56.
- D Schwartz. - *Méthodes statistiques à l'usage des médecins et des biologistes.* - 3^{ème} édition. Paris. Editions Flammarion Médecine-Sciences. 1986.
- Strum DP, May JH, Vargas LG. - *Modeling of uncertainty of surgical procedures times. Comparison of Log-normal and normal models.* Anesthesiology 2000 ; 92 : 1160 - 1167.
- Huguier M, Flahault A. - *Biostatistiques au quotidien.* 2^{ème} édition, Elsevier, Paris, 2003, 206 Pages.

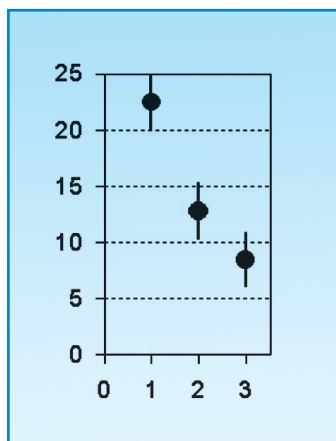


Figure 4.

INTERVALLE DE CONFIANCE :

L'intervalle de confiance à 95% de la moyenne est la fourchette dans laquelle se trouveraient statistiquement 95% des valeurs observées de la moyenne m d'un nombre infini d'échantillons de n individus, au sein d'une population de moyenne théorique μ [4].

$$IC_{95\%} = m \pm 1,96 \text{ ESM}$$

Pour les échantillons de petite taille où $n < 30$, la valeur 1,96 (≈ 2) doit être remplacée par une valeur supérieure, lue dans une table de distribution du t de Student en fonction du degré de liberté ($n - 1$) et du risque (nous verrons de quoi il s'agit ultérieurement) de première espèce α consenti [table dans 5].

Sur une représentation graphique, l'utilisation des intervalles de confiance permet d'une part au lecteur de se faire une idée précise de l'estimation des moyennes (Fig 4, détails dans [4]), et d'autre part, représente une comparaison statistique extrêmement intéressante en soi, car l'absence de recouvrement des intervalles de confiance signifie que la différence est statistiquement significative (entre série 1 et 2 de la figure 4), au contraire, le chevauchement signifie qu'il n'y a pas de différence (séries 2 et 3). Ici également, la valeur de l'intervalle de confiance (donc de l'incertitude) rend souvent absurde l'expression de décimales [7].

TRANSFORMATION z :

L'emploi de z est une habitude anglo-saxonne ; en France, le milieu médical préfère utiliser la table de l'écart-réduit, qui réalise exactement la même

procédure, c'est à dire la *normalisation de la distribution* observée par rapport à une distribution de Gauss définie par $\mu = 0$ et $\sigma = 1$ (coordonnées normales, centrées, réduites).

z exprime la distance à laquelle se situe une valeur x_i par rapport à la moyenne μ . C'est la différence entre cette valeur et la moyenne, rapportée à l'écart-type :

$$z = (x_i - \mu) / \sigma$$

La table de probabilité correspondante permet de rattacher à chaque z la probabilité d'observer une valeur $\geq x_i$ [3].

CONCLUSION

Ainsi, l'analyse des données brutes et, mieux, l'étude a priori des données qu'on pense obtenir, est le préalable indispensable à l'utilisation ultérieure des hypothèses et des tests statistiques appropriés. Cette série de « *digests* » de méthodologie statistique, destinée à vous permettre d'acquérir un esprit critique non seulement sur les données que vous récoltez, mais aussi sur certaines conclusions de la littérature [1], comprendra quatre autres opus : (2) comparaisons de moyennes, (3) tests non paramétriques, (4) utilisation des corrélations et (5), quelques notions d'épidémiologie. ■

Professeur Georges MION,
Docteurs Stéphane MÉRAT, Laurent GRASSER
Service d'Anesthésie - HIA Val de Grâce
Courriel : Georges.mion@club-internet.fr

Cet article a été soumis au comité de lecture d'Urgence Pratique et validé.