

Statistiques et kinésithérapie

Un abord simple avec les machines à calculer

G. PÉNINOU

M.C.M.K., membre du Cercle d'Études Kinésithérapiques, enseignant à l'E.F.O.M., 118 bis, rue de Javel, F 75015 Paris.

Cet article rédigé à la demande du comité de rédaction des Annales et impatientement attendu, a pour but de permettre aux lecteurs de se faire une idée de la valeur des « statistiques » présentées à l'appui de certains travaux, et de permettre aux auteurs éventuels de valider leurs collections de données, s'assurant que les résultats ne sont pas dus au hasard et exploitant au mieux les corrélations possibles.

Dissipons tout d'abord une confusion par trop répandue : le simple fait d'analyser des données par addition et soustraction ne constitue pas un exercice statistique, c'est tout bonnement une analyse arithmétique. Le fait de pouvoir dire : « nous avons eu plus de patients du type x cette année » relève de l'arithmétique. La capacité de dire « nous avons analysé les raisons, elles ne relèvent pas du hasard », devient une analyse statistique.

La confusion est malheureusement entretenue par les logiciels de gestion qui proposent une « fonction statistique » récapitulative du nombre d'événements similaires dans l'année (envoi par un médecin correspondant, paiement par la même mutuelle), alors qu'il s'agit plus simplement d'une fonction arithmétique d'addition par un appareil qui possède une mémoire d'éléphant : l'ordinateur.

Nous entendons trop fréquemment dire « ces résultats ont été validés par l'ordinateur » ; l'ordinateur ne valide rien, puisqu'il se borne à exécuter les ordres de la personne qui l'utilise. Cette dernière peut avoir fait un

usage rationnel de l'analyse statistique, ou non. Et si l'on a besoin de valider statistiquement point n'est besoin d'un appareil cher : un calculateur de prix moyen suffit, c'est ce que nous explique l'auteur.

Il ne s'agit pas ici d'expliquer ce que sont toutes les statistiques (ce serait bien long et il existe d'excellents ouvrages traitant du sujet) mais bien plus de permettre au plus grand nombre de faire, ou de comprendre, les statistiques les plus couramment employées en kinésithérapie.

Introduction

C'est un fait admis maintenant partout, les calculs statistiques envahissent tous les domaines. En kinésithérapie, c'est surtout dans la recherche qu'ils s'imposent.

Il n'y a pas de possibilité de validation technologique sérieuse sans une étude sur une population. Il n'est pas possible, à partir d'une étude de population, de ne pas envisager une étude statistique minimum. Il n'est plus actuellement envisageable de présenter ou d'assister à des communications orales ou écrites qui donnent des qualificatifs du genre « ça marche » ou du genre « j'ai des résultats » comme s'il n'y avait pas d'échecs, ni de résultats de moindre qualité.

Les statistiques remettent un peu d'ordre dans l'appréciation des séries, des groupes, des populations ou des résultats. Les machines à calculer modernes permettent au plus grand nombre,

sans grande connaissance mathématique, de résoudre la plupart des problèmes courant en kinésithérapie; mais encore faut-il pouvoir interpréter les chiffres affichés...

Éléments de base

**LOI NORMALE OU LOI DE LAPLACE
OU LOI DE GAUSS**

« La moyenne d'un échantillon extrait d'une population quelconque est distribuée selon une loi pratiquement normale, quand la taille de l'échantillon est suffisamment grande » (Lazar-Schwartz). Ce qui veut dire que toutes les populations importantes se distribuent suivant une répartition du même type, dispersée autour d'une moyenne.

Cela est représenté dans la courbe dite « de Gauss » (fig. 1). Par exemple, les résultats considérés comme « moyens » seraient les plus nombreux, les résultats excellents seraient eux très peu nombreux, et les échecs seraient également très peu nombreux. Entre ces extrêmes la répartition des résultats va suivre la courbe normale de Gauss, si la population est suffisamment importante (d'où la méfiance exprimée au vu de 5 cas).

Une étude sur quelques cas pouvant exceptionnellement se situer dans une partie de la courbe n'aurait aucune signification, car ce qui est important c'est :

- La moyenne exprimée suivant les cas en \bar{x} , ou μ , ou m , ou M .
- L'écart de cette moyenne, exprimée en écart-

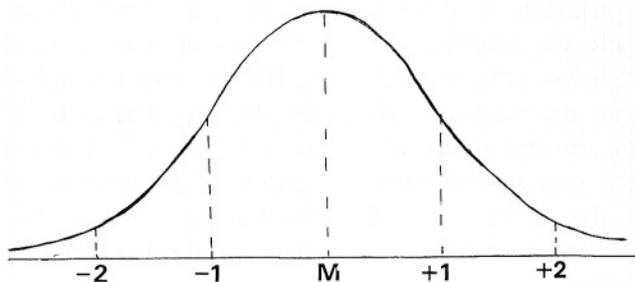


FIG. 1. - Courbe de Gauss.

- Entre -1σ et $+1\sigma$ on a 68,9 % des cas.
- Entre -2σ et $+2\sigma$ on a 95,5 % des cas.

Pour un même intervalle de classe c'est la distance qui fait ou non l'étalement de la courbe.

type suivant le cas en s , ou σ .

Cette moyenne et cette dispersion de la population sont les premiers calculs et parfois les seuls à faire en statistique.

**COMMENT FAIRE LES CALCULS \bar{x} ET s
AVEC UNE MACHINE A CALCULER**

- Soit par exemple une population de 6 personnes dont la taille et le poids sont :

Taille (cm) : 180 - 160 - 165 - 175 - 161 - 185
Poids (kg) : 75 - 60 - 70 - 75 - 59 - 80

- 1^{re} question : Quelle est la moyenne et l'écart-type des tailles ? exemple de calcul avec la machine H.P. 32E - l'article est basé sur le clavier (fig. 2).

On va utiliser la touche $\Sigma+$, qui non seulement additionne les chiffres dans une mémoire, mais les porte au carré dans une autre mémoire, tout en conservant dans une autre le

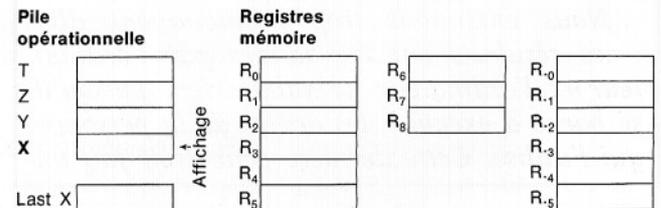
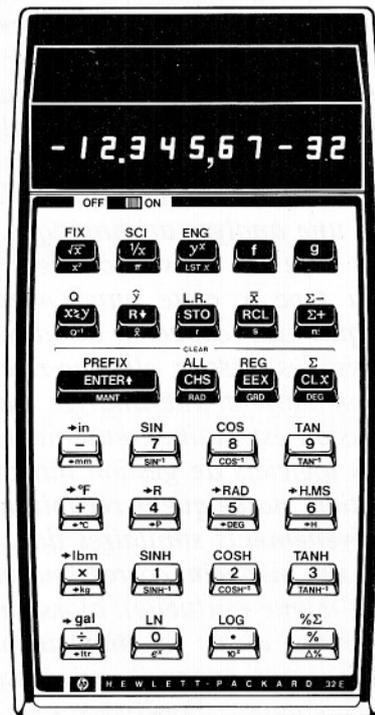


FIG. 2

nombre de chiffres rentrés.

Après avoir tapé le premier chiffre des tailles : 180, on appuie sur $\Sigma +$, alors apparaît le chiffre 1 qui veut dire que votre machine a enregistré le premier chiffre ; puis le deuxième chiffre : 160, on appuie sur $\Sigma +$, apparaît le chiffre 2. La machine fait aussitôt les différents calculs puis les place en mémoire, et ainsi de suite jusqu'à avoir rentré le dernier chiffre des tailles.

Ensuite il est possible par simple touche de connaître les résultats. Pour la moyenne il faut appuyer sur \bar{x} qui est sur la même touche que RCL ou S. Pour les différencier il existe la pré-touche \bar{f} en jaune, qui correspond à la série des opérations du dessus en jaune, et la pré-touche \bar{g} en bleu pour la partie des opérations du dessous ; soit : la taille de la population étudiée est

$$\bar{x} = \bar{f} \bar{x} = 171 \text{ cm}$$

$$s = \bar{g} s = 10,4881$$

La corrélation entre deux mesures

Le calcul de corrélation est exprimé par r , il cherche à discerner si 2 groupes de chiffres ont une relation entre eux. Par exemple : est-ce que, lorsque la taille de notre population de 6 sujets augmente, le poids augmente également ? On dit que la corrélation est parfaite si $r = 1$; qu'elle est nulle si $r = 0,0$.

CALCUL A L'AIDE DE LA CALCULATRICE HP 32E

Il suffit d'entrer les chiffres taille et poids de nos sujets en parallèle, c'est-à-dire 180 puis ENTER puis 75 puis $\Sigma +$ = 1
160 puis ENTER puis 60 puis $\Sigma +$ = 2
ainsi de suite

La population des tailles est stockée dans un registre Y. La population des poids, rentrée en dernier, est stockée dans le registre X.

Les mémoires enregistrant les chiffres, leur somme, leur moyenne, leur carré, leur produit.

La calculatrice nous donne \bar{x} des X, s des X, \bar{x} des Y, s des Y et r des X-Y

$$\text{Pour } \bar{x} X = \bar{f} \bar{x} = 69,8333 \text{ kg}$$

$$\bar{x} Y = \bar{x} > < y = 171 \text{ cm}$$

$$s X = \bar{g} s = 8,6120$$

$$s Y = \bar{x} > < y = 10,4881$$

$$r = \bar{g} \bar{r} = 0,9411$$

Ces opérations permettent de connaître la moyenne du poids : $x = 69,8333$ kg avec un écart-type de 8,6120 ; la moyenne de la taille : $y = 171$ cm avec un écart-type de 10,4881 (déjà calculé précédemment) ; le coefficient de corrélation $r = 0,9411$, un excellent coefficient qui indique que dans notre population, le poids et la taille croissent conjointement.

DROITE DE RÉGRESSION (fig. 3)

Dans beaucoup de publications, cette corrélation est objectivée par une droite moyenne dite droite de régression. Cette droite s'obtient à partir d'un graphique.

Sur l'axe des x nous portons les poids.

Sur l'axe des y nous portons les tailles.

Chaque croix représente un individu avec en abscisse et en ordonnée son poids et sa taille. Nous avons une suite de croix qui ne sont pas tout à fait alignées.

Pour avoir la droite moyenne des croix, il

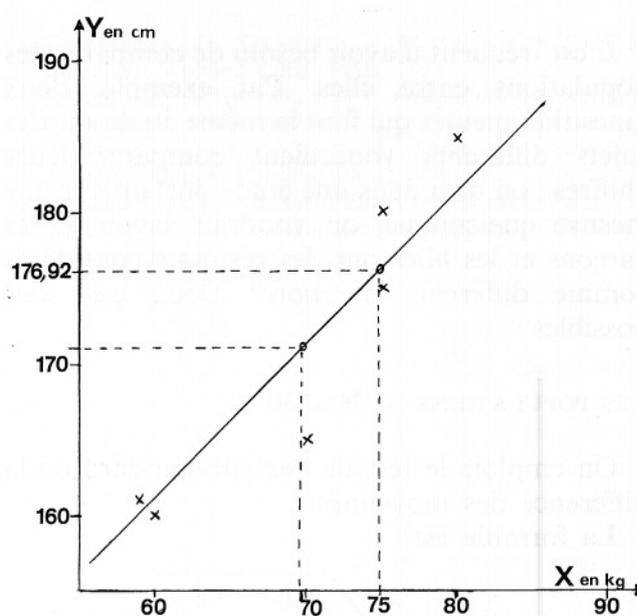


FIG. 3. - Droite de régression.

suffit d'avoir 2 points symbolisés ici par des petits cercles.

– Nous possédons un cercle qui sera le point représentant la moyenne poids et taille.

– Le deuxième point s'obtient en connaissant la variation de Y pour une variation de X ; c'est-à-dire dans notre exemple, pour une variation de 1 kg de combien varie la taille en cm ?

Cette variation s'obtient par la formule : $r \frac{\sigma Y}{\sigma X}$

Nous connaissons :

$$\begin{aligned} r &= 0,9411 \\ sY &= 10,4881 & \text{soit } r \frac{\sigma Y}{\sigma X} &= 1,1461 \\ sX &= 8,6120 \end{aligned}$$

Cela signifie que, quand le poids varie de 1 kg, la taille de notre population varie de 1,1461 cm, pour un poids X de 75 kg qui est à 5,1667 de la moyenne des poids qui était 69,8333 kg.

La taille Y va varier de $5,1667 \times 1,1461 = 5,9216$ cm par rapport à la moyenne. Soit la taille sera de $171 \text{ cm} + 5,9216 = 176,9216$; ce qui nous donne le deuxième point de la droite. Nous avons en fait épuré la formule qui est

$$Y = MY + r \frac{\sigma Y}{\sigma X} (X - MX)$$

Comparaisons de deux moyennes

Il est fréquent d'avoir besoin de comparer des populations entre elles. Par exemple, deux kinésithérapeutes qui font la même étude sur des sujets différents voudraient comparer leurs chiffres ; ou bien dans une étude portant sur une mesure quelconque, on voudrait savoir si les garçons et les filles ont des résultats considérés comme différents ou non ? Deux cas sont possibles :

LES POPULATIONS = $N > 30$

On emploie le test de l'erreur-standard de la différence des moyennes.

La formule est :

$$Sd = \sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}$$

Si $M_1 - M_2 \geq 2 Sd$, on dit que la différence est significative.

La calculatrice a permis de connaître les moyennes exprimées en M ou \bar{x} , elle a permis de connaître les écarts-types, exprimés en s ou σ , et on connaît le nombre d'individus pour chaque population.

La population 1 aura le nombre N_1 , moyenne M_1 , écart σ_1 .

La population 2 aura le nombre N_2 , moyenne M_2 , écart σ_2 .

LES POPULATIONS OU L'UNE D'ELLES = $N \leq 30$

On emploie le test « t » de Student-Fischer
La formule est la suivante

$$t = (M_1 - M_2) \sqrt{\frac{(N_1 + N_2 - 2) N_1 \cdot N_2}{(N_1 \sigma_1^2 + N_2 \sigma_2^2) (N_1 + N_2)}}$$

Tous ces chiffres peuvent être donnés à partir de la calculatrice. On obtient cependant un chiffre qui n'aurait aucune signification seul : on se reporte à une table à double entrée, qui est dans tous les manuels de statistiques ; on détermine le degré de liberté (d.d.l.) qui est égal à $N_1 + N_2 - 2$.

Si la valeur : « t » \geq à la colonne 0,05, ou 5 %, la différence est dite significative.

Si la valeur : « t » \geq à la colonne 0,01, la différence est dite hautement significative.

Si au contraire la valeur « t » est inférieure au chiffre de la colonne 0,05 on conclue que la différence est seulement due au hasard de l'échantillonnage.

CAS DES POPULATIONS APPARIÉES

Il y a appariement lorsque ce sont les mêmes individus qui forment les 2 séries de chiffres. Par exemple : on veut comparer la vitesse de récupération du quadriceps de tous les sujets d'un service ou d'un cabinet. On connaît la valeur de la R.M. de la 10^e séance et celle de la 15^e séance, est-ce que les différences observées sur l'ensemble des sujets sont dues au hasard ou est-ce un progrès significatif indiquant que

les critères technologiques utilisés sont corrects ?

Le test « t » sera également utilisé, mais le calcul sera un peu différent. On calcule la différence des performances : d, pour chaque sujet ; on obtient une suite de chiffres d1, d2, d3, etc. qui auront une moyenne md et un σ d

$$t = \frac{md}{\sqrt{\frac{\sigma d}{N}}} \text{ le nombre de d.d.l.} = N - 1$$

Puis on se reporte comme précédemment à la table « t »

Le test du χ^2 (khi-deux)

Ce test finalement assez peu utilisé, a pour but de savoir si une population suit la répartition de la loi normale ou non.

En effet, la plupart des personnes, qui étudient une population, font un histogramme et l'apparence de la forme de l'histogramme montre si l'on a un aspect rappelant la courbe de Gauss ou non. En cas de doute, le test un peu long s'exprime suivant la formule :

$$\chi^2 = \sum \frac{(fo - ft)}{ft}$$

fo = fréquence de classe observée

ft = fréquence théorique de classe

Pour obtenir le calcul un peu long de la ft, nous renvoyons le lecteur aux ouvrages de statistiques.

Conclusion

Nous avons voulu montrer que le calcul statistique était devenu, grâce aux machines à calculer spécialisées, un outil quotidien pour ceux qui veulent faire ou comprendre les calculs statistiques les plus courants.

Il suffit que la calculatrice possède une touche \bar{x} , s, et r pour que l'essentiel des calculs puissent être effectués même si l'on n'est pas spécialement mathématicien.

La plupart des publications médicales et kinésithérapiques se suffisent de ces calculs pour justifier les résultats.

Tableau des conventions

N, ou n	= Nombre de chiffres
\bar{x} , ou μ , ou m, ou M	= Moyenne
s, ou σ	= Ecart-type
r	= Corrélation
t	= Test de Student
sd	= Comparaison de 2 populations N1 et N2 > à 30
d.d.l.	= Degré de liberté
χ^2	= khi-deux = validation de la bonne répartition de la population

Bibliographie sommaire

1. FERMANIAN J., SALMOND. - *Mathématique P.C.E.M., statistique*. Armand Colin, Paris, 1974.
2. LAZAR Ph., SCHWARTZ D. - *Probabilités et statistique*. Flammarion, Paris, 1967.
3. MOREAU M., MATHIEU A. - *Statistique appliquée à l'expérimentation*. Eyrolles, Paris, 1979.
4. VANDERVAEL F. - *Biométrie Humaine*. Masson, Paris, 1964
5. VIALLET F. - *Statistique et Recherche appliquée*. Chotard et Associés, Paris, 1970.