

# JOURNAL DE LA SOCIÉTÉ STATISTIQUE DE PARIS

CHARLES RICHET

## **De la méthode statistique dans les sciences expérimentales**

*Journal de la société statistique de Paris*, tome 57 (1916), p. 466-473

[http://www.numdam.org/item?id=JSFS\\_1916\\_\\_57\\_\\_466\\_0](http://www.numdam.org/item?id=JSFS_1916__57__466_0)

© Société de statistique de Paris, 1916, tous droits réservés.

L'accès aux archives de la revue « Journal de la société statistique de Paris » (<http://publications-sfds.math.cnrs.fr/index.php/J-SFdS>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme  
Numérisation de documents anciens mathématiques  
<http://www.numdam.org/>

## II

### DE LA

### MÉTHODE STATISTIQUE DANS LES SCIENCES EXPÉRIMENTALES

On a beaucoup médité de la statistique; mais je suppose que c'est faute d'avoir suffisamment réfléchi; car, dans toute recherche scientifique, qu'on l'avoue ou qu'on le nie, on est forcé d'avoir recours à la statistique.

Autrement dit à la moyenne. Prendre une moyenne, c'est déjà faire de la statistique. Et il faut toujours prendre une moyenne, car sans cela on tomberait dans d'inexpiables erreurs.

Par exemple, je veux déterminer le poids atomique de l'argent : une première mesure me donne 108,2. Si je me contente de ce nombre, comment être assuré que je ne me suis pas prodigieusement trompé? *Una experientia, experientia nulla*. Tout à fait comme pour les témoignages, puisque aussi bien chaque expérience est un témoignage. Il faut donc de toute nécessité que je fasse une seconde expérience, au moins une autre. Si dans cette autre expérience je trouve 107,8, je puis supposer, si le même soin a été mis à l'une et à l'autre mesure, que le poids atomique de l'argent est la moyenne arithmétique entre 107,8 et 108,2, c'est-à-dire 108,0.

Plus j'accumulerai d'expériences (si elles sont aussi bien faites que le comportent mon habileté technique et l'état actuel de la science), plus j'approcherai du chiffre véritable. Si une troisième expérience me donne 107,6 et une quatrième 107,9, et une cinquième 108,0, et une sixième 107,8, la moyenne arithmétique des six sera 107,88, bien près d'être le nombre exact.

Il est vrai que je n'aurai pas ainsi éliminé une cause d'erreur systématique. La statistique ne supprime pas les causes d'erreur systématiques; car dans ces six expériences, par exemple, il se peut très bien qu'une parcelle d'eau, par suite de la déshydratation incomplète du chlorure d'argent due à

la défectuosité de mes appareils, soit restée, ce qui modifiera d'un centième environ, je suppose, pour chaque détermination, le chiffre obtenu. Qu'un autre observateur, plus sagace, élimine cette cause d'erreur, il trouvera 107,2, 106,8, 106,6, 106,9, 107,0, 106,8; mais finalement il sera toujours forcé de recourir à la statistique, c'est-à-dire à la moyenne, pour conclure à un seul chiffre au moyen des divers chiffres qu'il aura trouvés.

Il en est de même pour toutes les sciences, physiques et chimiques; *la méthode statistique n'est valable que lorsqu'on a éliminé les grosses causes d'erreur systématiques.*

Aussi faut-il être très prudent quand on veut prendre, pour aboutir à une mesure scientifique précise, les moyennes obtenues par divers savants; car enfin il en est, parmi ces savants, qui ont mal opéré, et qui, par conséquent, trouvent un chiffre constamment trop faible ou constamment trop fort. Ce serait folie de rectifier le bon chiffre, donné par un habile chimiste en lui adjoignant le mauvais chiffre, donné par un mauvais chimiste. J'aurai plus de confiance dans un seul document, entouré de précautions minutieuses, que dans six documents défectueux.

Oui! l'erreur serait grave si l'on croyait que la moyenne permet de conclure avec de mauvaises expériences.

Pour cela je prendrai un exemple évident.

L'excitation du nerf pneumogastrique arrête le cœur. Voilà un fait maintenant avéré et incontestable. Pourtant, il est possible, si j'emploie un appareil électrique défectueux, et si je fais douze expériences, que huit fois j'aie arrêté le cœur et que quatre fois je ne l'aie pas arrêté. Ne serait-il pas absurde de conclure que quatre fois sur douze le pneumogastrique n'est pas le nerf d'arrêt du cœur?

L'art de l'expérimentateur consiste à supprimer les variations dans les résultats et à obtenir constamment les mêmes effets quand il fait intervenir les mêmes causes.

On peut dire que la méthode statistique ne vaut pas pour la constatation des phénomènes mêmes, mais pour le *dosage* de ces phénomènes; elle est quantitative beaucoup plus que qualitative.

\* \* \*

Si la méthode statistique s'impose pour les recherches de chimie et de physique, elle s'impose avec plus de force encore pour les sciences physiologiques et médicales qui comportent tant d'incertitudes.

Voici par exemple un poison dont je veux étudier les effets. Je suppose qu'il s'agit du phosphore, et je prends précisément le cas d'une série d'expériences faites par moi récemment. La question que je m'étais posée était la suivante : Quelle est la dose toxique de phosphore pour un kilo de poids vif?

D'abord le poids d'un animal est une donnée assez inexacte. La balance nous indique bien le poids d'un chien. Mais ce chien a-t-il mangé ou est-il à jeun? Est-il gras ou est-il maigre? A-t-il uriné, ou a-t-il la vessie pleine? Est-il jeune, ou adulte, ou vieux? Est-ce un mâle ou une femelle?

Je veux bien qu'on élimine successivement toutes ces variations, et qu'on

prenne des chiens comparables, c'est-à-dire adultes et pas trop vieux, pas trop gras et pas trop maigres, mâles, ayant uriné et étant à jeun. Supposons tout cela. Pourtant les chiens divers ne seront pas exactement comparables. Un griffon, un terrier, un bulldog, un lévrier, un caniche, un terre-neuve, auront peut-être des réactions différentes et leur poids brut ne sera qu'une approximation imparfaite, car un kilo de caniche n'est pas du tout un kilo de lévrier. Alors je serai forcé de prendre la moyenne des diverses expériences, car jamais deux chiens ne sont identiques; ils se ressemblent bien moins que deux feuilles d'un même chêne; et pourtant les deux feuilles d'un même chêne sont toujours différentes.

Le seul procédé qui me permettra d'avoir un chiffre *équitable* sera de faire beaucoup d'expériences sur des chiens très divers, et de conclure d'après la moyenne.

Voici, pour fixer les idées, des chiffres se rapportant à la toxicité du phosphore pour les chiens (phosphure de zinc).

Dose de phosphure de zinc par kilo de poids vif		Dose de phosphure de zinc par kilo de poids vif	
—		—	
Milligrammes		Milligrammes	
0,45	Survie.	20,0	Survie.
0,45	—	24,0	—
0,9	—	24,0	—
1,4	—	24,0	—
2,1	—	24,0	—
5,0	—	25,0	—
5,0	—	25,0	—
6,2	—	25,0	—
9,3	—	25,0	—
10,0	—	25,0	Mort au 14 <sup>e</sup> jour.
10,0	—	28,0	— 2 <sup>e</sup> jour.
10,0	— (mais a été malade).	30,0	— 2 <sup>e</sup> jour.
15,0	—	30,0	— 2 <sup>e</sup> jour.
15,0	—	32,0	— 2 <sup>e</sup> jour.
15,0	Mort au 2 <sup>e</sup> jour.	35,0	— 1 <sup>er</sup> jour.
20,0	Survie.	45,0	— 10 <sup>e</sup> jour.
20,0	—	50,0	— 1 <sup>er</sup> jour.
20,0	—		

Tout de suite nous pouvons éliminer un chien qui est aberrant et qui, après la dose de 15 milligrammes, est mort au deuxième jour. Pourquoi? Je ne saurais le dire; mais ce que je sais, c'est qu'il fait exception et qu'il ne faut pas le compter dans notre moyenne. Si je n'avais pas voulu faire de statistique, et si la malchance m'avait fait faire cette unique expérience, j'aurais conclu (à tort) que la dose toxique de phosphure de zinc est inférieure à 15 milligrammes par kilo; ce qui est manifestement faux.

Peut-être aussi faut-il regarder comme aberrant le cas du chien qui n'est mort que le dixième jour, alors qu'il avait pris une dose très forte, 45 milligrammes.

En tout cas il est bien entendu que, même en multipliant les expériences, je ne pourrai jamais arriver à un chiffre absolu, à une seconde décimale dont je

sois certain. Toujours un moment viendra où il y aura hésitation, car les conditions pour les divers animaux ne pourront être les mêmes.

\* \* \*

Je prendrai un autre exemple de la méthode statistique dans les recherches que j'ai faites sur l'asphyxie; — je m'excuse de citer surtout mes expériences; ma justification est que je suis plus à même de les juger que de juger les expériences d'autrui.

Si l'on tue des chiens par submersion, on découvre bien vite qu'il faut au moins une minute quinze secondes de submersion pour les noyer. Tant qu'ils ne seront pas restés une minute quinze secondes sous l'eau, ils survivront. Il ne suffit pas d'une expérience pour le constater sûrement; il faut en faire au moins cinq ou six.

Au delà d'une minute quarante-cinq secondes, jamais ils ne survivent, quelles que soient les tentatives de respiration artificielle.

Par conséquent : 1<sup>o</sup> la durée d'une minute quinze secondes est toujours insuffisante à amener la mort;

2<sup>o</sup> La durée d'une minute quarante-cinq secondes est toujours suffisante à déterminer la mort.

Mais il reste une durée d'incertitude, une zone d'incertitude, que nous donne la statistique; car dans certains cas, à une minute vingt secondes, des animaux n'ont pas survécu, tandis qu'à une minute quarante secondes des animaux ont survécu. Pour construire cette zone d'incertitude, il faut une statistique, presque une statistique graphique, d'autant plus valable que le nombre des expériences aura été plus grand. Et toujours, malgré des expériences répétées, il y aura une zone incertaine, car les conditions ne pourront être toujours identiques; les animaux auront un sang plus ou moins chargé d'oxygène, feront des mouvements plus ou moins intenses, se débattront avec vigueur, ou resteront inertes; on pratiquera la respiration artificielle avec plus ou moins de précision; la quantité d'eau introduite dans le poumon sera plus ou moins considérable, et l'expulsion en sera plus ou moins rapide; toutes variations individuelles dont le détail est presque sans fin et qui ne permettent jamais de dire avec certitude : « Au bout d'une minute trente secondes de submersion, il y a toujours mort, ou il n'y a jamais mort définitive. »

Toutes les expériences de physiologie, ou à peu près toutes, comportent au point de vue quantitatif une variation, une incertitude.

Pour la toxicologie par exemple, il faut distinguer :

- α) la dose toujours mortelle;
- β) la dose quelquefois mortelle;
- γ) la dose jamais mortelle.

Il est vrai que la détermination de ces trois doses exige beaucoup de travail, et que par malheur beaucoup de savants se contentent de quelques expériences peu nombreuses. Elles ont beau être bien faites, ces peu nombreuses expériences, elles ne suffisent pas; car il faut se mettre à l'abri des écarts individuels, dès que l'on veut déterminer un chiffre ayant quelque généralité.

\* \* \*

Pour faire comprendre le sens des deux mots *qualitatif* et *quantitatif*, qui indiquent bien les limites dans lesquelles doit évoluer la statistique, je prendrai l'exemple suivant :

La mesure de l'acidité du lait qui a fermenté donne exactement l'intensité de la fermentation, et par conséquent, l'activité, la vitalité du ferment lactique.

Supposons que je place les laits à fermenter à des températures différentes. Alors j'aurai des écarts énormes. Par exemple je trouverai des chiffres tels que :

25  
4  
45  
8  
9

ce qui m'interdira toute moyenne, toute statistique.

La méthode statistique ne pourra s'appliquer que si les conditions sont à peu près semblables. Aussi, soumettant mes laits qui fermentent à une température identique, vais-je obtenir des chiffres très voisins.

25,8  
25,4  
25,7  
25,4  
25,6

} Moyenne : 25,58;  
Écart maximum : 0,4,

C'est pourquoi je pourrai dans ces conditions prendre la moyenne.

Cette moyenne, obtenue quand toutes les conditions connues de moi me paraissent identiques, sera une moyenne valable authentique; et le chiffre obtenu sera plus *vrai* que tel ou tel chiffre spécial.

Supposons — et c'est un bienfait de la méthode statistique — que je trouve ces chiffres trop éloignés l'un de l'autre. Alors j'examinerai avec un soin particulier pourquoi il y a 25,4 dans deux cas, et, dans un autre cas, 25,8, et je noterai que dans les deux cas où il y a eu 25,4 le lait fermentait dans des tubes un peu plus étroits que les autres, encore que la quantité de lait fût la même. Alors je referai une expérience en employant des flacons exactement semblables, et je trouverai, par exemple :

26,3  
26,2  
26,3  
26,4  
26,2

} Moyenne : 26,28;  
Écart maximum : 0,02.

J'aurai ainsi diminué l'écart entre mes expériences, et j'aurai pu démontrer que la forme du flacon exerce une influence sur la marche de la fermentation lactique.

C'est grâce à cette méthode statistique que j'ai pu, en 1902, découvrir le phénomène de l'anaphylaxie.

En étudiant sur les chiens la dose toxique de l'extrait glyciné des ten-

tacules d'actinies, j'ai trouvé les chiffres suivants, exprimant en centimètres cubes la dose mortelle pour un kilo de chien :

1,3	0,04
1,2	0,06
1,5	1,3
1,2	1,4
1,1	0,03

Et tout de suite on voit que ces chiffres hétéroclites ne peuvent fournir une moyenne.

Alors, en cherchant pourquoi, sur certains chiens, la dose toxique était si faible, j'ai vite reconnu que c'est parce que ces chiens étaient des animaux ayant été déjà injectés, par conséquent, étant dans un certain état d'hyper-sensibilité que j'ai appelé anaphylaxie.

Répétant ces expériences, et faisant deux catégories d'animaux : animaux neufs ou animaux ayant reçu une injection préalable, j'ai pu dresser le tableau suivant (schématique) :

**Dose toxique.**

Animaux neufs		Animaux précédemment injectés	
1,4	Moyenne : 1,31.	0,04	Moyenne : 0,042.
1,3		0,05	
1,3		0,04	
1,4		0,03	
1,3		0,05	
1,2		0,04	

Ainsi était démontré le phénomène de l'anaphylaxie, c'est-à-dire la sensibilisation par des doses antécédentes,

\* \* \*

Le dernier point que j'ai à traiter ici est un peu plus délicat et plus théorique. Mais je tâcherai de le rendre très accessible.

Malgré tous nos efforts, jamais, dans un dosage ou une mesure, nous ne pouvons atteindre l'absolue identité des chiffres. Si nous voulons aller jusqu'aux deuxièmes ou aux troisièmes décimales, toujours il y aura des divergences, car il ne peut y avoir identité entre deux expériences.

Prenons comme exemple la fermentation lactique. Voici les mêmes quantités de laitensemencées par le même ferment, à la même minute, placées dans des flacons rigoureusement identiques, à des températures absolument identiques, dosées aussi exactement que possible, par la même solution, et au même moment : tout de même je ne trouverai jamais identité ; j'aurai, pour obtenir la même coloration rose de l'indicateur (phtaléine du phénol) :

23,1	Moyenne : 23,0
23,0	
22,9	
23,0	
23,2	
22,9	

Pourquoi ces différences ? Pourquoi ne puis-je pas obtenir un écart nul ?

Nous sommes forcés de conclure qu'il y a des causes inconnues déterminant ces variations; causes qui sont sans doute multiples, et qui ne sont pas systématiques, ou du moins dont le système nous échappe complètement.

Appelons  $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\delta$  ces causes multiples de variations.

L'inconvénient en soi ne serait pas très grand, car il s'agit de quantités très faibles; et il est rare qu'on ait intérêt à avoir le chiffre de 23 ou celui de 23,2 pour une fermentation. Pourtant, il est un cas où cette cause de variation doit être corrigée, c'est lorsqu'on veut faire intervenir une influence connue, mais extrêmement faible, de l'ordre de  $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\delta$ . Appelons  $\eta$  cette cause de variation.

Je prends alors six séries d'expériences dans lesquelles il y a, mélangés aux autres, des flacons avec  $\eta$  (marqués d'une croix). J'aurai, par exemple :

23,1	22,9	23,2	23,1	23,0	22,9
23,0	23,1	23,0+	23,2	22,9	23,1+
22,9+	23,2	23,0	22,9	23,2	23,2
23,0	23,1+	22,9	23,0+	22,9+	23,0
23,2	22,9	23,1	23,1	23,2	22,9
22,9	23,1	22,9	23,2	23,1	23,2

Voici 36 expériences qui me donneront une moyenne générale de 23,044.

Mais, sur six de ces flacons, j'ai fait intervenir une influence (très faible)  $\eta$  (par exemple, l'addition d'une minuscule quantité de platine). Alors, prenant les six flacons avec platine, je trouve :

$$\left. \begin{array}{l} 22,9 \\ 23,1 \\ 23,0 \\ 23,0 \\ 22,9 \\ 23,1 \end{array} \right\} \text{Moyenne : } 23,0.$$

J'ai donc le droit de conclure que le platine, à dose minuscule, a exercé quelque influence, puisque la moyenne est 23,000 au lieu d'être 23,044.

Jamais je n'eusse pu découvrir cette action faible, si je n'avais pas multiplié les expériences; autrement dit, recouru à la méthode statistique.

Pour le dire en passant, c'est grâce à cette étude des fermentations lactiques que j'ai pu découvrir ce fait imprévu que des doses minuscules modifient la fermentation. Et quand je dis minuscules, je veux dire des chiffres effarants par leur petitesse. Dans un litre de lait, une quantité de vanadium est encore active, qui est égale à 0<sup>gr</sup> 000.000.001. Cela signifie qu'un centigramme de chlorure de vanadium peut agir sur le lait que produiraient en un jour un million de vaches.

\* \*

Un chiffre isolé ne signifie rien, car c'est un fait individuel. Or, la science peut se définir en disant que c'est la loi du général et non de l'individuel.

Voici un individu dont je mesure très exactement la température, avec un thermomètre merveilleux qui me donne le millième de degré, et je trouve 37° 128.



Eh bien ! ce chiffre n'est guère intéressant, si je n'en ai pas d'autre. En effet, pour peu que ce même individu se mette à me parler, avec quelque vivacité, sa température va s'élever à 37° 256. Si j'avais pris sa température à 4 heures du matin, pendant qu'il sommeillait profondément, j'eusse trouvé 36° 375. Et, s'il avait une forte fièvre, il m'eût donné 40° 344. Donc mon premier chiffre de 37° 128, si exact qu'il soit, n'a de valeur que s'il est encadré par d'autres chiffres.

Et alors sur des individus, de même âge, à la même heure, dans de mêmes conditions, je prendrai d'autres mesures et si je trouve (après 37° 128) :

37° 142

37 084

37 152

37 204

je pourrai tout de suite en déduire une moyenne, et le chiffre final (37° 142) sera plus *vrai* que chacun de ces chiffres considéré isolément.

Pour indiquer la température de l'homme, il faut un chiffre moyen.

La connaissance des moyennes est seule scientifique; les faits isolés, épars, même s'ils sont recueillis avec soin, n'appartiennent pas à la science. Ils ne signifient que par leurs groupements et leurs relations.

. \* \*

Et nous en concluons qu'il n'y a pas de plus grande ineptie, encore qu'elle soit répétée partout, que cette phrase stupide : « On fait dire aux chiffres ce que l'on veut ! » C'est faute de réflexion, faute d'intelligence, qu'on parle ainsi. Ne pas accepter ce que disent les chiffres, c'est tomber dans un fossé d'ignorance et de présomption, qui est sans fond. Et j'aime mieux dire avec l'Ecclésiaste : « *Omnia in numero et pondere.* »

Charles RICHET.

---