

LOGIQUE DES TESTS STATISTIQUES

M. CHASTRETTE(*), E. ROLETTO(**)

(*) Université de Lyon I,
Laboratoire de chimie organique physique
43 boulevard du 11 Novembre 1918
69622 Villeurbanne cedex

(**) Université de Turin

*Et nous sommes tenus, par toutes les règles de la justice
et de l'équité, de faire confiance à la nouvelle année
pour être bonne, jusqu'à ce qu'elle donne la preuve
qu'elle est indigne du crédit que nous lui accordons.*

Charles Dickens
dans Esquisses de BOZ : Personnages

I- LA LOGIQUE DES DECISIONS EN STATISTIQUE

Une des difficultés que présente la statistique pour les débutants est la démarche un peu déroutante suivie par le statisticien pour arriver à une conclusion.

Considérons un ensemble de patients atteints d'hypertension qui est divisé en deux groupes. Le premier groupe subit un traitement alors que le second est considéré comme groupe de contrôle et ne subit aucun traitement. A la fin de l'expérience on constate que la tension artérielle du groupe de contrôle est en moyenne plus élevée que celle du groupe expérimental.

La première démarche du statisticien est de mettre en concurrence deux explications des faits observés :

* La première explication, assez évidente, consiste à dire que l'effet observé est dû à ce que le traitement subi par le groupe expérimental est réellement efficace. Intuitivement cette explication semble d'autant plus crédible que la différence entre les deux groupes est plus marquée.

* La deuxième explication, qui va généralement à l'encontre des souhaits de l'expérimentateur, est que la différence observée n'est pas due au traitement mais, au contraire, aux effets cumulés de facteurs non contrôlés, ou en d'autres termes à l'effet du hasard. Après tout, personne ne s'attend à ce que les groupes soient strictement identiques et intuitivement cette explication est d'autant plus crédible que la différence entre les groupes est moins marquée.

La deuxième démarche du statisticien est de choisir de travailler sur la deuxième explication plutôt que sur la première. Il décide de considérer que la deuxième explication est vraie jusqu'à ce qu'il soit capable de la rejeter comme très peu probable. S'il la rejette finalement il accepte alors la première explication comme vraie.

Le côté troublant de cette démarche est que la conclusion "le traitement est considéré comme efficace" est tirée non pas d'une preuve positive de l'efficacité du traitement mais du rejet de la conclusion que "les effets observés peuvent être attribués au hasard".

Enfin il faut bien noter que le fait qu'on ne puisse pas rejeter la première explication n'implique pas qu'elle soit vraie. De la même manière le fait de ne pas prouver que l'accusé est coupable n'est pas la même chose que prouver son innocence. Dans le principe on présume l'accusé innocent jusqu'à ce qu'on ait pu prouver qu'il est coupable.

"Tests of significance, when used accurately, are capable of rejecting or invalidating hypotheses, in so far as these are contradicted by the data, but they are never capable of establishing them as certainly true". Ronald A. Fischer (1).

La troisième démarche du statisticien, qui fait toute la valeur de la conclusion finale est d'estimer le risque de se tromper en énonçant l'une ou l'autre des conclusions. L'estimation est faite à partir du calcul des probabilités d'apparition des effets observés.

En quelque sorte la statistique est un moyen de convertir des observations sur la taille d'un effet observé en probabilité d'apparition d'un tel effet par le seul jeu de hasard. Le test statistique nous fournit une estimation de la probabilité d'apparition par chance des effets observés. Si cette probabilité est très faible il y a très peu de chances que l'effet soit dû au hasard. On conçoit bien qu'il ne puisse pas exister un seul outil adapté à toutes les circonstances et qu'il faille choisir entre plusieurs de ces outils, ou tests statistiques, ceux qui sont les plus pertinents.

II- LA FORMALISATION DES HYPOTHESES

Les explications concurrentes mentionnées plus haut sont formalisées en suivant un certain nombre de conventions. L'idée fondamentale est qu'il est possible que les effets observés sur les variables dépendantes ne soient pas dus à l'influence des variables indépendantes mais à celle des seules variables non contrôlées.

II-1. L'HYPOTHESE NULLE H_0

Par convention on appelle hypothèse nulle, notée H_0 , l'hypothèse selon laquelle les variables indépendantes n'ont aucune influence sur les variables dépendantes.

Les tests statistiques convenables permettent d'évaluer la probabilité pour que H_0 soit vraie. Si l'on trouve que cette probabilité est très faible on rejette l'hypothèse nulle. Si la probabilité n'est pas suffisamment faible on continue à considérer H_0 comme vraie (ce qui n'implique pas nécessairement que H_0 soit réellement vraie).

II-2. L'HYPOTHESE ALTERNATIVE H_1

L'hypothèse qui intéresse le plus l'expérimentateur, à savoir que les variables qu'il manipule ont bien un effet sur les variables dépendantes est appelée hypothèse alternative H_1 . Elle est évidemment acceptée si H_0 est rejetée et inversement. H_1 a donc un rôle passif et dépend de H_0 . En fait si on y regarde de près il y a plusieurs espèces d'hypothèses alternatives.

L'hypothèse alternative générale H_1 , spécifie seulement que H_0 n'est pas vraie. Par exemple, si H_0 déclare que le traitement n'a pas d'effet, H_1 déclare que le traitement a bien un effet, sans préciser dans quel sens cet effet s'exerce.

L'hypothèse alternative spécifique H_1 , est plus ambitieuse que la précédente en ce sens qu'elle spécifie dans quel sens s'exerce l'effet. Par exemple H_1 ,

déclare que les patients traités présentent une amélioration significative de leur hypertension.

II-3. L'EVALUATION DE LA PROBABILITE POUR QUE H_0 SOIT VRAIE.

Une des difficultés souvent sous estimées est qu'on ne teste pas H_0 de la même manière selon que H_1 est générale ou spécifique. Il convient de bien faire la distinction. Ceci sera discuté sur des exemples simples, car la technique utilisée est étroitement liée au type de variables mises en jeu.

" *You can prove anything by statistics' is a common jibe. Its contrary is more nearly true - you can never prove anything by statistics. The statistician is dealing with the most complex cases of multiple causation. He may show that the facts are in accordance with this hypothesis or that. But it is quite another thing to show that all other possible hypotheses are excluded, and that the facts do not admit of any other interpretation than the particular one he may have in mind*". G.U. Yule (2).

II-4. LA FORMULATION DE LA DECISION

La décision est annoncée avec l'indication du risque d'erreur. On dira par exemple l'hypothèse H_1 est retenue au seuil de signification de 5 %. Ceci veut dire qu'il y a moins de 5 chances sur cent pour que le résultat observé soit dû au hasard. Les seuils de signification généralement retenus sont de :

5 % (0,05), 2,5% (0,025), 1 % (0,01),
0,5 % (0,005) et 0,1 % (0,001).

Conventions pour la formulation des conclusions :

> 5 %	non significatif
> 1 %	significatif
> 0,1 %	hautement significatif
< 0,1 %	extrêmement significatif

Le choix d'un seuil plutôt que d'un autre comme limite acceptable dépend du type de problème, de l'expérimentateur et aussi des risques encourus.

II-5. LES RISQUES D'ERREUR

Le risque de rejeter une hypothèse H_0 vraie en réalité est lié au seuil de probabilité retenu comme acceptable pour une influence significative. La probabilité d'accepter comme vraie une hypothèse H_0 fautive en réalité dépend aussi du seuil de probabilité retenu. Les quatre cas possibles sont représentés dans le tableau 1, avec les probabilités correspondantes :

Tableau 1

Règle suivie de décision au seuil choisi	en réalité	
	H_0 est vraie	H_0 est fautive
Ho rejetée	erreur du type I avec la proba= α	décision cor- recte prob= $1-\alpha$
Décision ----- Ho non rejetée	décision cor- recte prob= $1-\beta$	erreur du type II, prob = β

II-6. PUISSANCE D'UN TEST

Ce qui intéresse le plus le chercheur est de pouvoir rejeter H_0 quand elle est fautive car alors l'hypothèse alternative prend de la valeur. Un test est donc intéressant lorsqu'il est très sensible c'est à dire qu'il donne seulement de faibles taux d'erreurs du type II (accepter une hypothèse en réalité fautive).

Pour augmenter la sensibilité du test on peut diminuer la part de hasard en augmentant la taille de l'échantillon, en contrôlant mieux les variables, en faisant des mesures plus précises, etc... Mais une fois les données acquises on peut encore augmenter la sensibilité en rendant l'hypothèse H_1 aussi spécifique que possible.

La sensibilité définie plus haut est appelée par les statisticiens "puissance du test". Un test est d'autant plus puissant qu'il réduit les chances d'erreurs du type II. Il existe des divergences sur l'appréciation de la puissance de différents tests et le meilleur moyen d'accroître le résultat des tests est encore de collecter des données de qualité.

REFERENCES

- 1) Fischer R.A. Nature 1935, 136, p. 474
- 2) Yule G.U. "The Function of Scientific Method in Scientific Investigation", Industrial Fatigue Research Board Report : London, U.K., 1924, p 28