

<https://irem.unicaen.fr/spip.php?article180>



La statistique du chi-deux : son usage à partir de l'article de Doll et Hill sur l'association entre cancer et tabac (Mise à jour juin 2016).

BRITISH MEDICAL JOURNAL
LONDON SATURDAY SEPTEMBER 30 1950

SMOKING AND CARCINOMA OF THE LUNG
PRELIMINARY REPORT
BY
RICHARD DOLL, M.D., M.R.C.P.
Member of the Statistical Research Unit of the Medical Research Council
AND
A. BRADFORD HILL, PH.D., D.Sc.
Professor of Medical Statistics, London School of Hygiene and Tropical Medicine; Honorary Director of the Statistical
Research Unit of the Medical Research Council

Date de mise à jour : samedi 18 juin 2016

- Les Groupes - HISTOIRE DES SCIENCES - Groupe « Histoire des probabilités et de la statistique » -

Copyright © IREM de Caen Normandie - Tous droits réservés

Le document présenté ici est issu des travaux du groupe d'histoire de la statistique, StatMed, au sein du cercle d'histoire des sciences de l'IREM. Ces travaux ont donné lieu à une communication lors du colloque organisé par la Commission inter-IREM Épistémologie et Histoire des mathématiques au Mans dont le thème était : « Les Mathématiques et le réel : expériences, instruments, investigations », en mai 2015.

La question de l'introduction de méthodes numériques en médecine est au coeur des préoccupations de notre groupe de recherche. Cet article prolonge notre article de 2012, [Statistique inférentielle au fil de l'ouvrage de Jules Gavarret](#) où nous analysons l'ouvrage *Principes généraux de Statistique Médicale ou développement des règles qui doivent présider à son emploi*. Cet ouvrage, publié en 1840, nous semblait porter en germe de nombreux outils de la statistique inférentielle développés principalement dans le milieu scientifique anglo-saxon du début du 20e siècle et nous nous étions étonnés de l'absence de successeurs à Gavarret parmi les savants français entre 1850 et 1950, date à laquelle l'école d'épidémiologie française est apparue sous l'impulsion de Daniel Schwartz. Nous nous sommes alors intéressés aux outils utilisés dans les premières enquêtes épidémiologiques de taille importante et particulièrement, dans celle de R. Doll et A. B. Hill, *Smoking and Carcinoma of the Lung*, consacrée à la liaison entre tabac et cancer.

[JPEG - 25.2 ko] Karl Pearson	[JPEG - 8 ko] Austin Bradford Hill	[JPEG - 9.3 ko] Richard Doll
-------------------------------	------------------------------------	------------------------------

Ce travail a commencé en octobre 2012 et la question des relations entre cancer et statistique a rebondi récemment à propos de la polémique engendrée par la publication le 2 janvier 2015 de l'article de Cristian Tomasetti et Bert Vogelstein, *Variations in cancer risk among tissues can be explained by the number of stem divisions*, dans la revue *Science*. Le titre de l'étude est assez clair : *La variation dans les risques de cancer entre différents organes peut être expliquée par le nombre de divisions des cellules souches*. La confusion s'est accrue avec une phrase de l'article : « Ces résultats suggèrent qu'un tiers seulement de la variation dans les risques de cancer entre différents organes peut être attribuée à des facteurs environnementaux ou des prédispositions héréditaires. La majorité est due à la malchance. ». Cette phrase déformée, simplifiée d'abord par les éditeurs de la revue, puis par des journalistes spécialisés est devenue : « deux tiers des cancers sont dus à la malchance, et non à la pollution, aux industries agroalimentaires, chimiques ou cigarettières ».

Nous nous intéressons dans un premier temps à la nature possible de cette relation et aux deux grands types d'études qui continuent d'être utilisées pour tenter de la quantifier. Nous rappelons ensuite les méthodes statistiques de comparaison de proportions, puis, après avoir donné les diverses expressions que peut prendre une statistique dite du chi-deux, nous présentons l'usage qu'en font Doll et Hill et les conclusions tirées de leur étude. Cet article est complété par quelques rappels sur les prolongements que cet article a eu, les polémiques qu'il a suscitées et par les controverses que provoquent aujourd'hui d'autres études portant sur la santé publique.

Dans le texte publié ici, nous avons inséré des activités. Les activités sont de types et de niveaux très variables. Certaines approfondissent le texte de Doll et Hill en proposant de démontrer ou de faire effectivement les calculs indiqués par nos auteurs. D'autres tentent d'illustrer des situations ou des problèmes mis en évidence dans leur texte. Ces activités peuvent donc être utilisées par un lecteur attentif pour prolonger sa lecture. Elles peuvent aussi donner lieu à des séances de formation continue pour des enseignants de mathématiques (niveau lycée et/ou premières années post-bac). Elles pourront enfin être la source d'activités avec des élèves ou des étudiants.

En fin d'article, nous indiquons la présentation actuelle de la statistique du chi-deux et les différentes utilisations qui peuvent en être faites, illustrées par des exemples tirés, entre autres, de l'ouvrage de Gavarret.

Le fichier de l'article peut être téléchargé ci-dessous.

Fichier mis à jour en juin 2016 : ajout d'activités et de leurs solutions.

Nous accueillerons avec plaisir toute réaction, critique, correction ... de la part des lecteurs, auprès de l'IREM.

Nous vous souhaitons une bonne lecture.

Le groupe d'histoire de la statistique de l'IREM

Jacques Faisant, Denis Lanier, Jean Lejeune, Rémy Morello, Didier Trotoux [PDF - 2.2 Mo](#) **La statistique du chi-deux : son usage à partir de l'article de Doll et Hill sur l'association entre cancer et tabac.**