Journées Nationales APMEP LAON 2015 Compte-rendu de l'atelier L30

« Mathématiques et tabac : une histoire conjointe au XX° siècle et aujourd'hui »

(Travail du groupe « Histoire des probabilités et de la statistique », IREM de Caen - Normandie,
Didier Trotoux, Rémy Morello, Jean Lejeune, Denis Lanier, Jacques Faisant)

19 octobre 2015

Première partie
Le compte rendu

Cet atelier[1] n'a pas attiré un grand nombre de congressites : seulement cinq.

De plus, la description chronologique de l'évolution conjointe, et de la diffusion du tabac, et des méthodes statistiques qui ont finalement permis de mettre en évidence sa nocivité, a surpris les participants.

Bien sûr, l'utilisation en classe de cet élément d'histoire contemporaine n'est pas immédiate ; j'espère néanmoins avoir apporté des informations utiles.

(On peut retrouver ces informations et d'autres sur la page WEB « JN2015 » indiquée en note ci-dessous.)

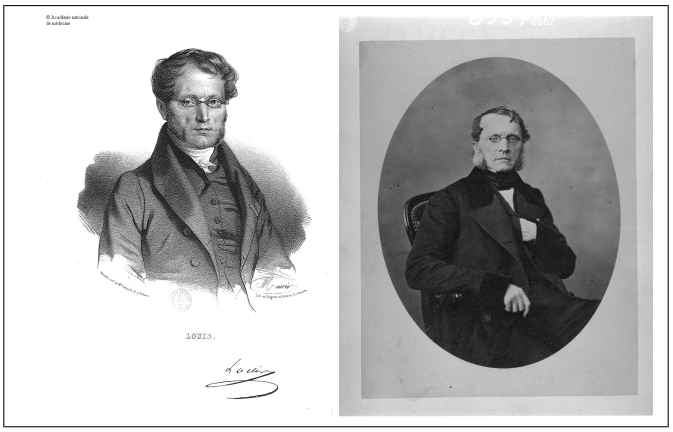
Jacques Faisant, professeur retraité

(Michel Fréchet l'a fait remarquer : beaucoup de retraités figurent parmi les animateurs d'ateliers ...)

Deuxième partie
Le schéma de la communication

1 Les protagonistes

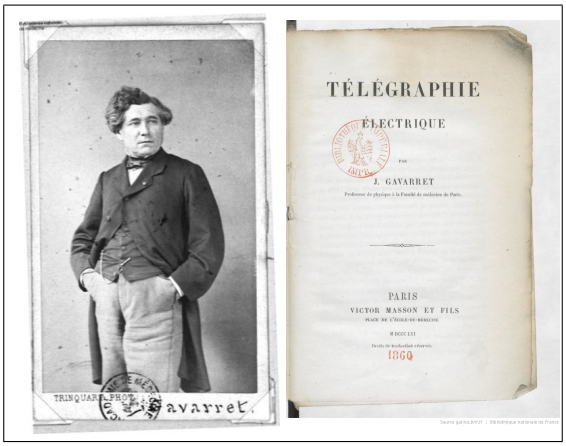
Le Docteur Louis



Le Docteur Pierre Charles Alexandre Louis (1787-1872), clinicien
- médecin à l'Hôpital de la Pitié, puis à l'Hôtel Dieu.
- célèbre grâce à sa méthode numérique en médecine et à l'usage des statistiques (← BNF)

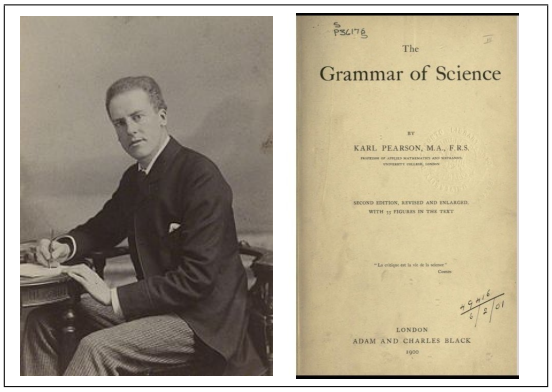
[1] http ://www.math.unicaen.fr/irem/spip.php ?rubrique43 ; http ://jacques.faisant.pagesperso-orange.fr/JN2015/

Le Docteur Gavarret



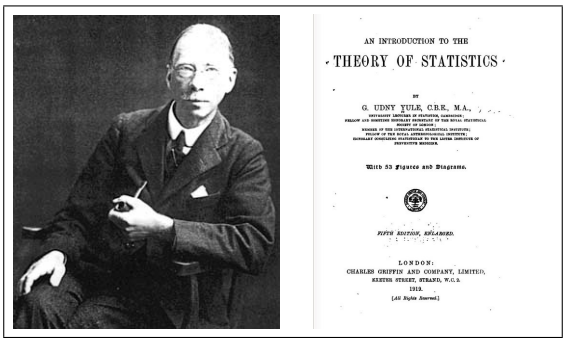
Jules Louis Dominique Gavarret (1809-1890),
– Diplômé de l'École Polytechnique, artilleur
– Docteur en médecine, professeur de physique médicale et inspecteur général de l'Instruction publique
– Renommé pour avoir développé la méthodologie statistique en médecine (← fr.wikipedia)

Karl Pearson



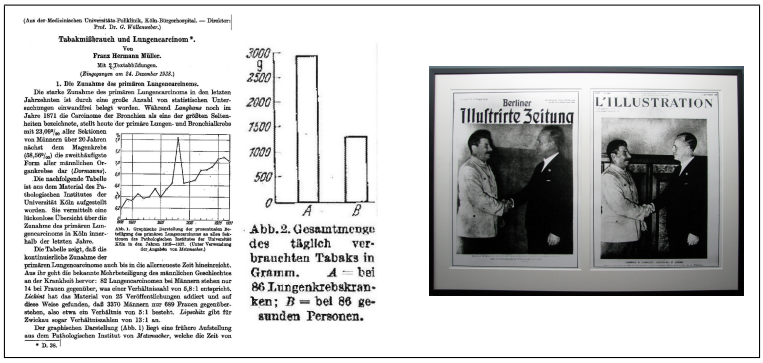
Karl Pearson (1857-1936)
– Historien, germaniste et mathématicien (il fut « Third–Wrangler » au Tripos de mathématiques de Cambridge)
– Enseignant à l'University College, Londres, 1883
– Inventeur du coefficient de corrélation linéaire et du **test du chi-deux** (← en.wikipedia)

George Udny Yule



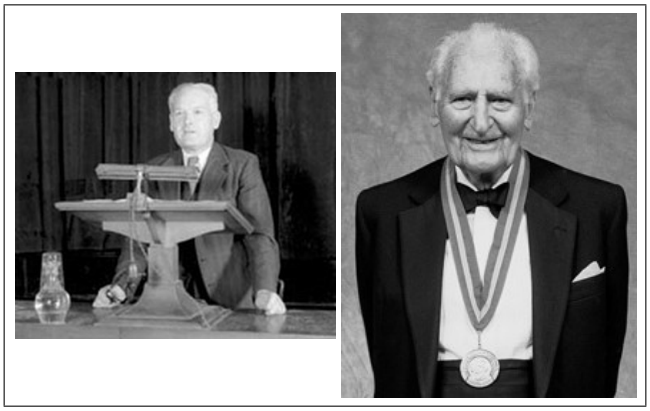
George Udny Yule (1871–1951)
– admis à l'âge 16 ans à l'University College de Londres en sciences de l'ingénieur
– préparateur pour Karl Pearson, puis assistant (1893-1899)
– professeur à l'Université de Cambridge (1912)
– Ce statisticien écossais est notamment à l'origine de la notion de **rapport de cotes** (ou « *odds ratio* ») (← en.wikipedia)

Franz Hermann Müller



Franz Hermann Müller
– médecin à Cologne
– a publié en 1939 une enquête approfondie sur la relation entre l'usage du tabac et le cancer du poumon
– du fait des crimes nazis, son travail n'a pas été connu internationalement et est tombé dans l'oubli (← de.wikipedia)

Austin Bradford Hill et Richard Doll



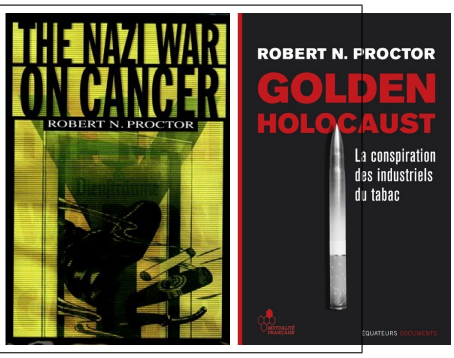
Austin Bradford Hill (1897–1991) Richard Doll (1912-2005)

Les héros de l'histoire !

Philip Morris (et les autres)

Johnny Roventini (1910-1998)

Robert Neel Proctor



Robert Neel Proctor

- historien des sciences,
- Pennsylvania State University puis Stanford University
- premier historien à témoigner (en 1999) contre l'industrie du tabac
- créateur de l'agnotologie
  ou étude de l'ignorance ou du doute induits via la publication de données scientifiques inexactes ou trompeuses (<- en.wikipedia)

Une question très actuelle ...

CANCER ETIOLOGY

Variation in cancer risk among tissues can be explained by the number of stem cell divisions

Cristian Tomasetti[1]* and Bert Vogelstein[2]*

Some tissue types give rise to human cancers millions of times more often than other tissue types. Although this has been recognized for more than a century, it has never been explained. Here, we show that the lifetime risk of cancers of many different types is strongly correlated (0.81) with the total number of divisions of the normal self-renewing cells maintaining that tissue's homeostasis. These results suggest that only a third of the variation in cancer risk among tissues is attributable to environmental factors or inherited predispositions. The majority is due to "bad luck," that is, random mutations arising during DNA replication in normal, noncancerous stem cells. This is important not only for understanding the disease but also for designing strategies to limit the mortality it causes.

Extreme variation in cancer incidence across different tissues is well known; for example, the lifetime risk of being diagnosed with cancer is 6.9% for lung, 1.08% for thyroid, 0.6% for brain and the rest of the nervous system, 0.003% for pelvic bone and 0.00072% for laryngeal cartilage (*1*–*3*). Some of these differences are associated with well-known risk factors such as smoking, alcohol use, ultraviolet light, or human papilloma virus (HPV) (*4*, *5*), but this applies only to specific populations exposed to potent mutagens or viruses. And such exposures cannot explain why cancer risk in tissues within the alimentary tract can differ by as much as a factor of 24 [esophagus (0.51%), large intestine (4.82%), small intestine (0.20%), and stomach (0.86%)] (*3*). Moreover, cancers of the small intestinal epithelium are three times less common than brain tumors (*3*), even though small intestinal epithelial cells are exposed to much higher levels of environmental mutagens than are cells within the brain, which are protected by the blood-brain barrier.

Another well-studied contributor to cancer is inherited genetic variation. However, only 5 to 10% of cancers have a heritable component (*6*–*8*), and even when hereditary factors in predisposed individuals can be identified, the way in which these factors contribute to differences in cancer incidences among different organs is obscure. For example, the same, inherited mutant *APC* gene is responsible for both the predisposition to colorectal and small intestinal cancers

[1]Division of Biostatistics and Bioinformatics, Department of Oncology, Sidney Kimmel Cancer Center, Johns Hopkins University School of Medicine and Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, 550 North Broadway, Baltimore, MD 21205, USA. [2]Ludwig Center for Cancer Genetics and Therapeutics and Howard Hughes Medical Institute, Johns Hopkins Kimmel Cancer Center, 1650 Orleans Street, Baltimore, MD 21205, USA.
*Corresponding author. E-mail: ctomasetti@jhu.edu (C.T.); vogelbe@jhmi.edu (B.V.)

Corrected 23 January 2015; see full text.

- C. Tomasetti and B. Volgestein in Science, 02/01/2015 : These results suggest that only a third of the variation in cancer risk among tissues is attributable to environmental factors or inherited predisposition. The majority is due to « bad luck ».
- A. Thibaud-Mony in Le Monde, 07/01/2015 : Non le cancer n'est pas le fruit du hasard.
- S. Cabut in Le Monde, 02/01/2015 : Une chose est sûre, les sciences mathématiques occupent une place croissante en cancérologie.

2 La preuve par les mathématiques : le test du chi-deux d'homogénéité

2.1 La recherche d'une relation entre le tabagisme et le cancer du poumon

La recherche d'une relation entre le tabagisme et le cancer du poumon

Comment les lois de probabilité calculées par Karl Pearson en 1900, de fonctions de répartition :

$$x \mapsto 1 - \mathrm{P}$$

avec

$$\mathrm{P} = \sqrt{\frac{2}{\pi}} \int_{\chi}^{\infty} e^{-\frac{1}{2}\chi^2} d\chi$$
$$+ \sqrt{\frac{2}{\pi}} e^{-\frac{1}{2}\chi^2} \left(\frac{\chi}{1} + \frac{\chi^3}{1.3} + \frac{\chi^5}{1.3.5} + \cdots + \frac{\chi^{n-2}}{1.3.5 \ldots n-2} \right)$$

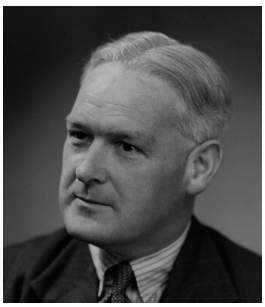
ainsi que le test du chi-deux dont elles sont la base, ont-elles été utilisés en 1950 par Richard Doll et Austin Bradford Hill pour inférer, à partir d'une enquête de grande ampleur, une relation entre le tabagisme et le cancer du poumon ?

2.2 Qui étaient Doll et Hill ?

Austin Bradford Hill

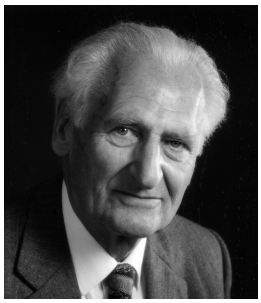
– **Sir Austin Bradford Hill** (1897-1991) engagé dans des études médicales, les abandonne pour raison de santé et fait carrière dans les statistiques médicales au Medical Research Council puis à l'École d'Hygiène et de Médecine Tropicale de Londres. En 1937, il publie *Principles of Medical Statistics*, livre de référence en statistique médicale qui aura 11 rééditions successives (12e édition en 1991).

men is 67·38 inches and the mean height of a group of 1304 Scotchmen is 68·61 inches. Are Scotchmen on the average taller than Englishmen or is the difference merely due to chance, inherent in sampling ? The standard error of the mean is, it has been shown, $\sigma/\sqrt{n}$, where σ is the standard



Richard Shaboe Doll

– **Sir Richard Doll** (1912-2005) abandonne ses études de mathématiques pour se tourner vers la médecine, études à la Faculté de l'hôpital St Thomas, King's College de Londres. Travaille au Middlesex Central Hospital avec l'unité de recherche du Medical Research Council où il rencontre A.B. Hill. Publication avec A.B. Hill dans le British Medical Journal le 30 septembre 1950 d'une étude portant sur le tabagisme et sa relation avec le cancer du poumon qui marque l'histoire scientifique médicale au XXe siècle.



2.3 Étaient-ils les premiers ?

Étaient-ils les premiers ?

Non, deux exemples

– **Franz Hermann Müller**, médecin allemand, compare en 1939, 86 cas de cancer du poumon masculins à 86 témoins sains, du même âge que les cas atteints de cancer du poumon. Constat : les victimes du cancer du poumon avaient une probabilité six fois plus élevée d'être d'« extrêmement gros fumeurs ».

– **Ernest L. Wynder**, étudiant à l'Université de Washington et **Evarts A. Graham** son professeur, publient en mai 1950 un article suite à une étude portant sur 605 patients atteints du cancer du poumon et 780 témoins, indemnes de cette maladie.

2.4 Motivations de l'étude de Doll et Hill

Accroissement, en Angleterre et au Pays de Galles, **du taux de mortalité dû au cancer du poumon**.

Entre 1927 et 1947, le nombre de décès est effectivement passé de 612 à 9287, accroissement sans commune mesure avec l'accroissement de la population.

Facteurs explicatifs envisagés, entre autres :

– Amélioration des techniques de diagnostic

– Pollution atmosphérique générale provenant des échappements des automobiles, des poussières de revêtements des routes, des usines à gaz, des usines industrielles et des centrales à charbon

– Augmentation de la consommation de tabac

◇ R. Doll et A. B. Hill s'orientent vers la recherche d'une **relation entre consommation de tabac et cancer du poumon**.

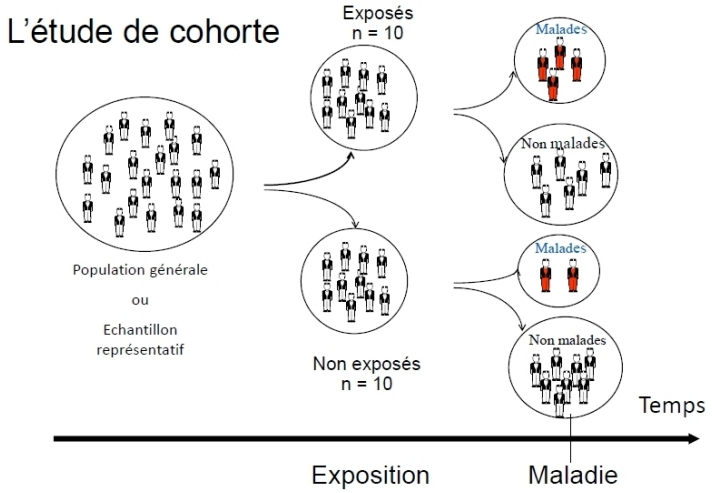
2.5 Formalisation mathématique des relations entre deux événements

Relation logique ?

- **Quelles «relations» existe-t-il entre les événements :**
  E_1 : «Être et/ou avoir été fumeur», E_0 : «Ni avoir été fumeur, ni l'être» (Exposition ou non au facteur de risque)
  et
  M_1 : «Contracter le cancer du poumon», M_0 : «Ne pas contracter le cancer du poumon» (Malade ou non)
- $E_1 \Rightarrow M_1$: FAUX
- $E_0 \Rightarrow M_0$: FAUX
- ◇ Au-delà de la seule logique, usage du calcul des probabilités et de la statistique inférentielle

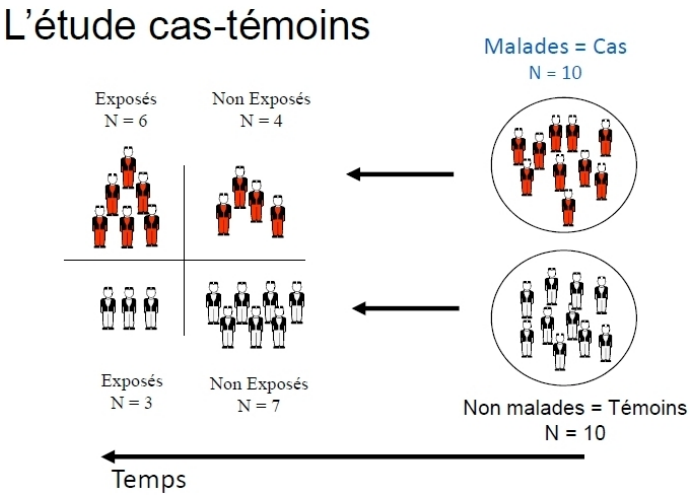
2.6 L'étude de cohortes

- **Première situation** : M_1 aléatoire (M_0 idem). **Étude de cohorte**s.
  $p_1 = \mathbb{P}_{E_1}(M_1)$ = probabilité pour un fumeur de contracter le cancer du poumon.
  Risque r_1 chez les épidémiologistes.
  $p_0 = \mathbb{P}_{E_0}(M_1)$= probabilité pour un non-fumeur de contracter le cancer du poumon.
  Risque r_0 chez les épidémiologistes.
  Problème : comparer p_0 et p_1.
  Si $p_0 = p_1$, E_1 n'est pas un facteur de risque



2.7 L'étude cas-témoins

- **Deuxième situation** : E_1 aléatoire (E_0 idem). **Étude cas-témoins.**
  $p'_1 = \mathbb{P}_{M_1}(E_1)$= probabilité pour quelqu'un qui a le cancer du poumon d'être fumeur
  $p'_0 = \mathbb{P}_{M_0}(E_1)$= probabilité pour quelqu'un qui n'a pas le cancer du poumon d'être fumeur
  Problème : comparer p'_0 et p'_1.



2.8 Comparaison de probabilités

Étude de cohortes

- L_1 sujets se sont trouvés exposés au facteur de risque. Parmi ceux-ci, a ont été malades et b n'ont pas été malades.
- L_0 sujets ne se sont pas trouvés exposés au facteur de risque. Parmi ceux-ci, c ont été malades et d n'ont pas été malades.

	Malades	Non malades	Total
Exposés	a	b	L_1 (fixé)
Non exposés	c	d	L_0 (fixé)
Total	$a+c$	$b+d$	T

Exemple fictif	Malades	Non malades	Total
Exposés	125	375	500
Non exposés	100	400	500
Total	225	775	1000

La « méthode numérique » popularisée par le docteur Louis (1787-1872) et l'objection de Gavarret

Comparer simplement les proportions observées : puisque $f_1 = \frac{a}{L_1} = \frac{125}{500} = 0,25$, la fréquence d'apparition de la maladie chez les exposés est supérieure à $f_0 = \frac{c}{L_0} = \frac{100}{500} = 0,20$, fréquence d'apparition de la maladie chez les non-exposés, l'exposition au facteur de risque accroit la probabilité de contracter la maladie.

Objection de Gavarret (1809-1890) en 1840 : il faut comparer $d = |f_1 - f_0|$, appelée maintenant variation absolue des risques, avec la « limite *compatible* avec l'*invariabilité des causes* », notée l. Conclusion :

- si $d > l$, l'exposition au facteur de risque modifie la probabilité de contracter la maladie
- si $d \leq l$, « tout porte à croire » que l'exposition au facteur de risque ne modifie pas la probabilité de contracter la maladie.

Calcul de l suivant Gavarret (qui utilise une formule due à Poisson) :

$$l = 2\sqrt{2}\sqrt{\frac{f_0(1-f_0)}{L_0} + \frac{f_1(1-f_1)}{L_1}} = 0,07456$$

Conclusion paraphrasant Gavarret : puisque la différence des fréquences $d = 0,25 - 0,20 = 0,05$ est inférieure à cette limite, « tout porte à croire » que l'exposition au facteur de risque ne modifie pas la probabilité de contracter la maladie.

2.9 Le test du chi-deux de Pearson

Le test du chi-deux de Pearson, une autre façon de comparer des proportions

Utilisation intensive par Doll et Hill de ce test.

- Sous l'hypothèse H_0 étudiée : probabilité d'être malade quand on est exposé = probabilité d'être malade quand on n'est pas exposé, il existe une probabilité inconnue p^* avec $p^* = p_0 = p_1$,
- p^* estimée par $f^* = \frac{a+c}{T} = \frac{125+100}{1000} = 0,225$.
- $1 - p^*$ estimée par $1 - f^* = \frac{b+d}{T} = 0,775$

Le test du chi-deux de Pearson utilise les estimations des **effectifs espérés** *sous l'hypothèse* H_0 :

$$a' = L_1 \times \frac{a+c}{T},\ c' = L_0 \times \frac{a+c}{T}$$

$$b' = L_1 \times \frac{b+d}{T},\ d' = L_0 \times \frac{b+d}{T}$$

Effectifs espérés	Malades	Non malades	Total
Exposés	a'	b'	L_1
Non exposés	c'	d'	L_0
Total	$a+c$	$b+d$	T

Effectifs espérés, même exemple fictif	Malades	Non malades	Total
Exposés	112,5	387,5	500
Non exposés	112,5	387,5	500
Total	225	775	1000

2.10 La statistique du chi-deux de Pearson

$$\chi^2_{\text{obs}} = \frac{(a-a')^2}{a'} + \frac{(b-b')^2}{b'} + \frac{(c-c')^2}{c'} + \frac{(d-d')^2}{d'}$$

– Avec le risque $\alpha = 5\%$, rejet de $H_0 : p_1 = p_0$ si et seulement si $\chi^2_{\text{obs}} > 3,84$.

– *Application à l'exemple* : $\chi^2_{\text{obs}} = 3,5842 < 3,84$. On ne rejette pas H_0.

Une formule simple, déjà donnée par Hill dans les *Principles of Medical Statistics* (1937).

Données	Malades	Non malades
Exposés	a	b
Non exposés	c	d

$$\chi^2_{\text{obs}} = \frac{(ad-bc)^2(a+b+c+d)}{(a+b)(c+d)(a+c)(b+d)}$$

2.11 Origine de la statistique du chi-deux

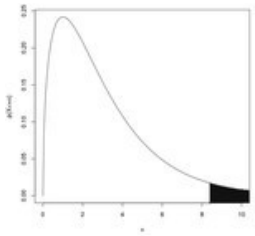
En 1900, Karl Pearson a proposé d'utiliser χ^2, une mesure de « distance », dite du chi-deux, entre les données observées lors d'une expérience aléatoire et les données **espérées** si certaines hypothèses probabilistes sont vraies.

Il donne explicitement les formules qui permettent de calculer la probabilité P (appelée aujourd'hui « p-value » ou « degré de signification ») que la distance du chi-deux soit supérieure ou égale au nombre χ^2_{obs}, calculé à partir des résultats de l'expérience. Elle dépend d'un nombre n appelé
« nombre de degrés de liberté ».

Si P est faible (resp. forte) , il conclut à la non adéquation (resp. à l'adéquation) des données observées avec les hypothèses probabilistes préalablement émises.

2.12 Table du chi-deux

R. A. Fisher (1890-1962) a établi en 1922, après K. Pearson, une table (reprise par Hill dans son livre) qui donne, les quantiles $\chi^2_P(n)$, nombres définis par :
$\mathbb{P}(Y \geq \chi^2_P(n)) = P$ où Y suit une loi du chi-deux à n degrés de liberté.



PRINCIPLES OF MEDICAL STATISTICS 309

TABLE OF χ^2 (*contd.*)

·50	·30	·20	·10	·05	·02	·01	n
·455	1·074	1·642	2·706	3·841	5·412	6·635	1
1·386	2·408	3·219	4·605	5·991	7·824	9·210	2
2·366	3·665	4·642	6·251	7·815	9·837	11·341	3
3·357	4·878	5·989	7·779	9·488	11·668	13·277	4
4·351	6·064	7·289	9·236	11·070	13·388	15·086	5
5·348	7·231	8·558	10·645	12·592	15·033	16·812	6
6·346	8·383	9·803	12·017	14·067	16·622	18·475	7
7·344	9·524	11·030	13·362	15·507	18·168	20·090	8
8·343	10·656	12·242	14·684	16·919	19·679	21·666	9
9·342	11·781	13·442	15·987	18·307	21·161	23·209	10
10·341	12·899	14·631	17·275	19·675	22·618	24·725	11
11·340	14·011	15·812	18·549	21·026	24·054	26·217	12
12·340	15·119	16·985	19·812	22·362	25·472	27·688	13
13·339	16·222	18·151	21·064	23·685	26·873	29·141	14
14·339	17·322	19·311	22·307	24·996	28·259	30·578	15
15·338	18·418	20·465	23·542	26·296	29·633	32·000	16
16·338	19·511	21·615	24·769	27·587	30·995	33·409	17
17·338	20·601	22·760	25·989	28·869	32·346	34·805	18
18·338	21·689	23·900	27·204	30·144	33·687	36·191	19
19·337	22·775	25·038	28·412	31·410	35·020	37·566	20

2.13 Application du test à l'étude cas-témoins

	Cas	Témoins	Total
Exposés	a	b	a+b
Non exposés	c	d	$c+d$
Total	C_1 (fixé)	C_0 (fixé)	T

L'hypothèse H_0 étudiée est : probabilité d'être exposé quand on est malade = probabilité d'être exposé quand on n'est pas malade.

On note p'_1 (resp. p'_0) la probabilité d'avoir été exposé au facteur de risque pour les « cas » (les malades) (resp. pour les témoins (les non-malades)).

$$\chi^2_{\text{obs}} = \frac{(ad-bc)^2(a+b+c+d)}{(a+b)(c+d)(a+c)(b+d)}$$

Avec le risque $\alpha = 5\%$, rejet de H_0 : $p'_1 = p'_0$ si et seulement si $\chi^2_{\text{obs}} > 3,84$.

2.14 L'enquête de Doll et Hill

La collecte des données (1948-1949)

- **20 hôpitaux londoniens**, enquête sur les malades atteints d'une des quatre pathologies suivantes : cancer du poumon, de l'estomac, du côlon ou du rectum.
- **Pour chaque malade atteint** du cancer du poumon (**cas**), **recherche** dans le même hôpital ou un hôpital voisin, **d'un patient** de même sexe et dans la même tranche d'âge (parmi 5 tranches d'âge) **non atteint** du cancer du poumon appelé par les auteurs «contrôle» (**témoin**).
- Au final, 709 patients atteints du cancer du poumon (cas) et 709 patients non atteints du cancer du poumon (**témoins**).

Le résultat premier

TABLE IV.—*Proportion of Smokers and Non-smokers in Lung-carcinoma Patients and in Control Patients with Diseases Other Than Cancer*

Disease Group	No. of Non-smokers	No. of Smokers	Probability Test
Males:			
Lung-carcinoma patients (649)	2 (0·3%)	647	P (exact method) = 0·00000064
Control patients with diseases other than cancer (649) ..	27 (4·2%)	622	
Females:			
Lung-carcinoma patients (60)	19 (31·7%)	41	$\chi^2 = 5{\cdot}76$; n = 1 0·01 < P < 0·02
Control patients with diseases other than cancer (60) ..	32 (53·3%)	28	

Rappel : $\chi^2_{\text{obs}} = \frac{(ad-bc)^2(a+b+c+d)}{(a+b)(c+d)(a+c)(b+d)}$

Calcul de χ^2_{obs} pour les hommes :

$\chi^2_{\text{obs}} = \frac{(2.622-27.647)^2(2+647+27+622)}{(2+647)(27+622)(2+27)(647+622)}$

$\chi^2_{\text{obs}} \cong 22.04 > 3,84$; $P \cong 2,7.10^{-6}$

Calcul de χ^2_{obs} pour les femmes :

$\chi^2_{\text{obs}} = \frac{(19.28-32.41)^2(19+41+32+28)}{(19+41)(32+28)(19+32)(41+28)}$

$\chi^2_{\text{obs}} \cong 5.76 > 3,84$; $P \cong 0,016$

NB : sont considérés comme fumeurs les patients qui avaient fumé au moins une cigarette par jour pendant au moins un an.

Conclusion : au risque $\alpha = 5\%$, on affirme l'**existence d'une association** entre le fait d'avoir été fumeur et le fait d'être atteint d'un cancer du poumon :

pour un homme comme pour une femme, la probabilité d'avoir été fumeur quand on souffre d'un cancer du poumon est différente de celle d'avoir été fumeur quand on n'en souffre pas.

2.15 Homogénéité des groupes

Doll et Hill se sont préoccupés de la cohérence de leur étude ; par construction, les deux groupes avaient des compositions identiques concernant le sexe et âge. En était-il de même pour la catégorie sociale, ou encore le type de lieu de résidence ?

Test du chi-deux pour la catégorie sociale (hommes seulement).

Pour le calcul dans le cas de deux lignes et trois colonnes, la formule de Pearson est, avec $(2-1).(3-1) = 2$ degrés de liberté :

$\chi^2_{\text{obs}} = \frac{(a-a')^2}{a'} + \frac{(b-b')^2}{b'} + \frac{(c-c')^2}{c'} + \frac{(d-d')^2}{d'} + \frac{(e-e')^2}{e'} + \frac{(f-f')^2}{f'}$

avec $a' = L_1 \times \frac{a+d}{T}$, $d' = L_0 \times \frac{a+d}{T}$, etc ...

Comparison Between Lung-carcinoma Patients and Non-cancer Patients Selected as Controls TABLE II.

Social Class (Registrar-General's Categories, Men Only)	No. of Lung-carcinoma Patients	No. of Non-cancer Patients
I and II ..	77	87
III ..	388	396
IV and V ..	184	166
All classes ..	649	649
Place of residence		
County of London ..	330	377
Outer London ..	203	231
Other county borough ..	23	16
Urban district ..	95	54
Rural district ..	43	27
Abroad or in Services ..	15	4
Total (M + F) ..	709	709

Effectifs observés :

Catégorie sociale	I et II	III	IV et V	Total
Patients atteints du cancer du poumon	a=77	b=388	c=184	L_1=649
Patients indemnes de cancer du poumon	d=87	e=396	f=166	L_0=649
Total	a+d=164	b+e=784	c+f=350	1298

Effectifs espérés :

Catégorie sociale	I et II	III	IV et V	Total
Patients atteints du cancer du poumon	a'=82	b'=392	c'=175	649
Patients indemnes de cancer du poumon	d'=82	e'=392	f'=175	649
Total	164	784	350	1298

Ainsi $\chi^2_{obs} = 2 \cdot \frac{5^2}{82} + 2 \cdot \frac{4^2}{392} + 2 \cdot \frac{9^2}{175} \cong 1,62 < 5,991$; P $\cong$ 0.445

- Donc rien n'empêche de penser que le groupe des cas et le groupe des témoins sont homogènes quant à leur classe sociale.
- Par contre, on peut affirmer, au risque $\alpha = 5\%$, que le groupe des cas et le groupe des témoins ne sont pas homogènes quant à leur lieu de résidence.

TABLE OF χ^2 (*contd.*)

·50	·30	·20	·10	·05	·02	·01	n
·455	1·074	1·642	2·706	3·841	5·412	6·635	1
1·386	2·408	3·219	4·605	5·991	7·824	9·210	2
2·366	3·665	4·642	6·251	7·815	9·837	11·341	3
3·357	4·878	5·989	7·779	9·488	11·668	13·277	4
4·351	6·064	7·289	9·236	11·070	13·388	15·086	5
5·348	7·231	8·558	10·645	12·592	15·033	16·812	6
6·346	8·383	9·803	12·017	14·067	16·622	18·475	7
7·344	9·524	11·030	13·362	15·507	18·168	20·090	8
8·343	10·656	12·242	14·684	16·919	19·679	21·666	9
9·342	11·781	13·442	15·987	18·307	21·161	23·209	10

The difference in social class distribution is small and is no more than might easily be due to chance (χ^2=1.61 ; n=2 ; 0.30<P<0.50). The difference in place of residence is, however, large (χ^2=31.49 ; n=5 ; P<0.001), and Table II shows that a higher proportion of the lung patients were resident outside London at the time of their admission to hospital.

2.16 Approfondissement de l'étude

Dans le tableau V, une distinction est faite entre les «fumeurs lourds» et les «fumeurs légers».

TABLE V.—*Most Recent Amount of Tobacco* Consumed Regularly by Smokers Before the Onset of Present Illness; Lung-carcinoma Patients and Control Patients with Diseases Other Than Cancer*

Disease Group	No. Smoking Daily					Probability Test
	1 Cig.–*	5 Cigs.–	15 Cigs.–	25 Cigs.–	50 Cigs.+	
Males: Lung-carcinoma patients (647)	33 (5·1%)	250 (38·6%)	196 (30·3%)	136 (21·0%)	32 (5·0%)	χ^2=36·95; n=4; P<0·001
Control patients with diseases other than cancer (622)	55 (8·8%)	293 (47·1%)	190 (30·5%)	71 (11·4%)	13 (2·1%)	
Females: Lung-carcinoma patients (41)	7 (17·1%)	19 (46·3%)	9 (22·0%)	6 (14·6%)	0 (0·0%)	χ^2=5·72; n=2; 0·05<P<0·10 (Women smoking 15 or more cigarettes a day grouped together)
Control patients with diseases other than cancer (28)	12 (42·9%)	10 (35·7%)	6 (21·4%)	0 (0·0%)	0 (0·0%)	

* Ounces of tobacco have been expressed as being equivalent to so many cigarettes. There is 1 oz. of tobacco in 26·5 normal-size cigarettes, so that the conversion factor has been taken as: 1 oz. of tobacco a week = 4 cigarettes a day.

P $\cong 1,8.10^{-7}$

P $\cong 0,057$

2.17 Autres critères caractérisant la consommation de tabac et leurs liens avec le cancer du poumon

Tester, contre son contraire, H_0 : absence de différence de comportement entre les personnes atteintes de cancer du poumon et les autres malades quant à :

- la consommation maximale journalière dans une période antérieure (H : P < 0,001 et F : 0,02 < P < 0,05)
- l'estimation du nombre total de cigarettes fumées depuis le début de la consommation de tabac (H : $P < 0,001$ et F : $0,001 < P < 0,01$)
- l'âge de début de la consommation de tabac (H et F : $0,30 < P < 0,50$)
- la durée de consommation de tabac (H et F : $0,05 < P < 0,10$)
- la durée de l'arrêt de consommation de tabac (H et F : $0,01 < P < 0,02$)
- au mode de consommation : cigarettes ou pipe (0,01 < P < 0,02)

† au fait d'avaler la fumée (H et F : 0,02 < P < 0,05) , les personnes atteintes d'un cancer du poumon avalent moins la fumée que les autres !

□ **Mais les études suivantes ont montré le contraire !**

N'oublions pas qu'employer les méthodes de la statistique inférentielle sert à contenir le risque d'erreur, mais pas à le supprimer.

2.18 L'influence de l'étude

Un tournant décisif au niveau scientifique :

– **Utilisation systématique du test du chi-deux**, pour inférer une association entre le tabagisme et le cancer du poumon en testant aussi l'homogénéité des deux groupes de patients pour vérifier l'absence d'un éventuel biais.

Socialement, une début de réussite :

– « ... travail scrupuleusement argumenté (et rigoureux sur le plan mathématique) de Doll et Hill ... les preuves étaient solides, cohérentes, sans ambiguïté. » (*Robert Proctor*).

– Cependant, l'étude cas-témoins venait seulement d'apparaître et le public, soumis aux campagnes de dénigrement de l'industrie du tabac, pouvait avoir des difficultés à en accepter les mécanismes.

– Doll et Hill décident alors de se lancer dans une étude de cohorte à grande échelle, la « British Doctors Study ».

2.19 La « British Doctors Study »

– Nouveau travail de **Doll et Hill**.
Étude d'une cohorte de 40000 médecins de Grande Bretagne faisant l'objet de publications en 1954, 1956 et 1964 dont les conclusions vont dans le même sens que le rapport de 1950. Cette cohorte a été suivie jusqu'en 2002.

– *Une conclusion majeure de l'étude est que le tabagisme diminue l'espérance de vie de 10 ans, et que plus de 50 % des fumeurs meurent d'une maladie connue pour être liée au tabagisme.*

– *Cependant, ceux qui arrêtent de fumer à 60, 50, 40 ou 30 ans augmentent leur espérance de vie de, respectivement, environ 3, 6, 9 ou 10 ans. Les fumeurs ayant arrêté de fumer avant l'âge de 30 ans ont donc la même espérance de vie que les non-fumeurs.*

3 Une approche plus récente de l'étude d'un éventuel facteur de risque

3.1 Les risques et les cotes : des indicateurs épidémiologiques

Cohortes	Malades	Non malades	Total
Exposés	a	b	L_1
Non exposés	c	d	L_0
Total	$a+c$	$b+d$	T

– **Les risques à comparer :**
$r_1 = \mathbb{P}_{\text{Exposé}}(\text{Malade})$,
estimé par $R_1 = \frac{a}{L_1}$ et
$r_0 = \mathbb{P}_{\text{Non exposé}}(\text{Malade})$
estimé par $R_0 = \frac{c}{L_0}$.

– **Le rapport des risques ou risque relatif :**
$rr = \frac{\mathbb{P}_{\text{Exposé}}(\text{Malade})}{\mathbb{P}_{\text{Non exposé}}(\text{Malade})} = \frac{r_1}{r_0}$, estimé par $RR = \frac{R_1}{R_0}$.
Le facteur de « risque » n'a pas d'influence sur la survenue de la maladie ssi $rr = 1$.

– **La cote de la maladie sous exposition au facteur de risque**
$o_{E_1} = \frac{\mathbb{P}_{\text{Exposé}}(\text{Malade})}{\mathbb{P}_{\text{Exposé}}(\text{Non Malade})} = \frac{r_1}{1-r_1}$, estimée par $\frac{\frac{a}{L_1}}{1-\frac{a}{L_1}} = \frac{a}{b}$.

– **La cote de la maladie sous non-exposition au facteur de risque**
$o_{E_0} = \frac{\mathbb{P}_{\text{Non Exposé}}(\text{Malade})}{\mathbb{P}_{\text{Non Exposé}}(\text{Non Malade})} = \frac{r_0}{1-r_0}$ estimée par $\frac{\frac{c}{L_0}}{1-\frac{c}{L_0}} = \frac{c}{d}$

Cas-témoins	Malades	Non malades	Total
Exposés	a	b	$a+b$
Non exposés	c	d	$c+d$
Total	C_1	C_0	T

– **La cote de l'exposition au facteur de risque chez les cas**
(Malades) : $o_{M_1} = \frac{\mathbb{P}_{\text{Malade}}(\text{Exposé})}{\mathbb{P}_{\text{Malade}}(\text{Non Exposé})}$
estimée par $\frac{\frac{a}{C_1}}{1-\frac{a}{C_1}} = \frac{a}{c}$

– **La cote de l'exposition au facteur de risque chez les témoins**
(Non malades) $o_{M_0} = \frac{\mathbb{P}_{\text{Non Malade}}(\text{Exposé})}{\mathbb{P}_{\text{Non Malade}}(\text{Non Exposé})}$ estimée par $\frac{\frac{b}{C_0}}{1-\frac{b}{C_0}} = \frac{b}{d}$

3.2 Les rapports de cotes ou odds ratios

– **Rapport des cotes de la maladie** (Odds ratio de la maladie)
$or_M = \frac{o_{E_1}}{o_{E_0}} = \frac{\frac{\mathbb{P}_{\text{Exposé}}(\text{Malade})}{\mathbb{P}_{\text{Exposé}}(\text{Non Malade})}}{\frac{\mathbb{P}_{\text{Non Exposé}}(\text{Malade})}{\mathbb{P}_{\text{Non Exposé}}(\text{Non Malade})}}$ estimé par $OR_M = \frac{\frac{\frac{a}{L_1}}{1-\frac{a}{L_1}}}{\frac{\frac{c}{L_0}}{1-\frac{c}{L_0}}} = \frac{\frac{a}{b}}{\frac{c}{d}} = \frac{a\times d}{b\times c}$

– **Rapport des cotes de l'exposition au facteur de risque** (Odds ratio de l'exposition au facteur de risque)
$or_E = \frac{o_{M_1}}{o_{M_0}} = \frac{\frac{\mathbb{P}_{\text{Malade}}(\text{Exposé})}{\mathbb{P}_{\text{Malade}}(\text{Non Exposé})}}{\frac{\mathbb{P}_{\text{Non Malade}}(\text{Exposé})}{\mathbb{P}_{\text{Non Malade}}(\text{Non Exposé})}}$ estimé par $OR_E = \frac{\frac{\frac{a}{C_1}}{1-\frac{a}{C_1}}}{\frac{\frac{b}{C_0}}{1-\frac{b}{C_0}}} = \frac{\frac{a}{c}}{\frac{b}{d}} = \frac{a\times d}{b\times c}$

Propriété

Si r_1 et r_0 « petits », alors $or_M = \frac{\frac{r_1}{1-r_1}}{\frac{r_0}{1-r_0}} \cong rr$

Les définitions de Yule (1900)

$Q = \frac{ad-bc}{ad+bc}$
$\kappa = \frac{1-Q}{1+Q} = \frac{bc}{ad} = \frac{1}{or_M}$

	Malades	Non malades
Exposés	a	b
Non exposés	c	d

3.3 Interprétation du rapport des cotes de la maladie

Résultat essentiel : $x \mapsto \frac{x}{1-x}$ définit une bijection strictement croissante de $[0\,;\,1[$ vers $[0\,;\,+\infty[$. Donc :

- si or_M vaut 1, l'exposition ou la non-exposition au facteur supposé à risque ne change pas la probabilité de voir survenir la pathologie considérée.
- Si $or_M = n$ avec $n > 1$, alors les sujets exposés au facteur supposé à risque ont plus de « chances » (*n fois plus de « chances » si r_1 et r_0 sont petits*) de voir survenir la pathologie que les sujets non exposés à ce même facteur : le facteur est à risque délétère.
- Si $or_M = n$ avec $n < 1$, alors les sujets non exposés au facteur de risque ont plus de « chances » (*$\frac{1}{n}$ fois plus de « chances » si r_1 et r_0 sont petits*) de voir survenir la pathologie que les sujets exposés à ce même facteur : le facteur est protecteur de la maladie.

3.4 Quel rapport de cotes ?

La définition de la probabilité conditionnelle nous donne :

$$\mathbb{P}_E(M) = \frac{\mathbb{P}_M(E).\mathbb{P}(M)}{\mathbb{P}(E)} \text{ et } \mathbb{P}_{\overline{E}}(M) = \frac{\mathbb{P}_M(\overline{E}).\mathbb{P}(M)}{\mathbb{P}(\overline{E})}$$

donc

$$\frac{\mathbb{P}_E(M)}{\mathbb{P}_{\overline{E}}(M)} = \frac{\mathbb{P}_M(E)}{\mathbb{P}_M(\overline{E})}.\frac{\mathbb{P}(\overline{E})}{\mathbb{P}(E)}$$

L'odds ratio de la maladie est $or_M = \frac{\frac{\mathbb{P}_E(M)}{\mathbb{P}_E(\overline{M})}}{\frac{\mathbb{P}_{\overline{E}}(M)}{\mathbb{P}_{\overline{E}}(\overline{M})}} = \frac{\mathbb{P}_E(M).\mathbb{P}_{\overline{E}}(\overline{M})}{\mathbb{P}_{\overline{E}}(M).\mathbb{P}_E(\overline{M})}$.

Comme plus haut,

$$\frac{\mathbb{P}_{\overline{E}}(\overline{M})}{\mathbb{P}_E(\overline{M})} = \frac{\mathbb{P}_{\overline{M}}(\overline{E})}{\mathbb{P}_{\overline{M}}(E)}.\frac{\mathbb{P}(E)}{\mathbb{P}(\overline{E})}$$

Par conséquent, l'odds ratio de la maladie est tel que

$$or_M = \frac{\mathbb{P}_E(M)}{\mathbb{P}_{\overline{E}}(M)}.\frac{\mathbb{P}_{\overline{E}}(\overline{M})}{\mathbb{P}_E(\overline{M})} = \frac{\mathbb{P}_M(E)}{\mathbb{P}_M(\overline{E})}.\frac{\mathbb{P}(\overline{E})}{\mathbb{P}(E)}.\frac{\mathbb{P}_{\overline{M}}(\overline{E})}{\mathbb{P}_{\overline{M}}(E)}.\frac{\mathbb{P}(E)}{\mathbb{P}(\overline{E})} = \frac{\mathbb{P}_M(E).\mathbb{P}_{\overline{M}}(\overline{E})}{\mathbb{P}_M(\overline{E}).\mathbb{P}_{\overline{M}}(E)} = or_E$$

Il n'y a qu'un seul rapport de cotes !

- **L'odds ratio de la maladie est donc égal à l'odd ratio de l'exposition**.
- Il n'y a qu'un seul type de rapport de cotes, qui peut être estimé par $\frac{a.d}{c.b}$ **même si on ne dispose que d'une étude cas-témoins.**
  Ce type d'étude, dans le cas où r_1 et r_0 sont proches de 0, permet donc d'obtenir, via l'odds ratio, une approximation du risque relatif rr

Conclusion

Bien que r_1 et r_0 ne puissent pas être estimés à partir d'une étude cas-témoins, d'après ce qui précède, on peut quand même estimer **approximativement** le rapport des risques rr en calculant OR_E à partir du tableau de l'enquête cas-témoins, si les risques r_1 et r_0 sont petits.

3.5 Rapports de cotes : test, intervalle de confiance

- Test du rapport de cotes
  rappel : $\chi^2_{\text{obs}} = \frac{(ad-bc)^2(a+b+c+d)}{(a+b)(c+d)(a+c)(b+d)}$
  Si $\chi^2_{\text{obs}} > 3,84$, on peut rejeter, au risque $\alpha = 5\,\%$, l'hypothèse $or_E = 1$ et/ou $or_M = 1$.
  En effet, cette hypothèse équivaut à l'absence d'influence du facteur sur la survenue de la maladie.
- Construction de l'intervalle de confiance à 95% pour rr par la méthode de Woolf[2] :
  $\left[OR.e^{-1,96\sqrt{\frac{1}{a}+\frac{1}{b}+\frac{1}{c}+\frac{1}{d}}}\,;\,OR.e^{1,96\sqrt{\frac{1}{a}+\frac{1}{b}+\frac{1}{c}+\frac{1}{d}}}\right]$

[2] Voir la démonstration correspondante en annexe.

3.6 Application à des données de Doll et Hill

Application à des données de Doll et Hill

	Cas	Témoins	Total
Fumeurs	688	650	1338
Non fumeurs	21	59	80
Total	709	709	1418

$\chi^2_{\text{obs}} = 19,13 > 3,84$; $P = 0,000012 < 0,05$

$OR = \frac{688.59}{650.21} = 2,97$

Méthode de Woolf : $[1,79 ; 4,95]$

D'après ces calculs,

- **le rapport de cotes observé et donc, approximativement, le rapport observé des risques, sont significativement différents de 1** (deux méthodes),
- **le risque de survenue d'un cancer du poumon est entre 2 et 5 fois plus élevé chez les fumeurs que chez les non-fumeurs**.
  (Ceci suppose que cette maladie soit « suffisamment » rare, même chez les fumeurs.)

4 Conclusion

Cette histoire met en lumière deux aspects importants.

Le succès du travail de Doll et Hill n'est pas seulement dû au résultat obtenu, mais aussi à l'extraordinaire luxe de précautions prises par les auteurs pour essayer d'éliminer tous les biais et par l'utilisation intensive d'un outil mathématique, le test du chi-deux d'homogénéité. C'est l'outil statistique qui donne ici la rigueur et l'objectivité (et pas seulement son apparence).

Enfin, on peut remarquer que Doll et Hill ne parlent pas de causalité mais d'association (ou de corrélation) (thème actuel : livre de Jacques Attali, Peut-on prévoir l'avenir). L'étude statistique permet de quantifier des relations, des augmentations (ou diminutions) des risques. Mais le risque reste toujours du domaine de la probabilité.

Même si on évite en épidémiologie de parler de nombre de chances (d'avoir une maladie !) ou de cas favorables (au décès d'un malade !), on parle toujours en terme de probabilité, donc de hasard. La statistique médicale n'a pas pour objet d'éliminer toute forme de hasard pour découvrir les vrais causes (sociales, environnementales, industriels, ...) d'une maladie, mais d'encadrer ce hasard pour mieux le comprendre et éventuellement le conjurer. Que des scientifiques et des journalistes spécialisés tombent encore dans ces travers, comme on l'a vu en introduction, cela montre qu'il y a beaucoup à faire dans l'enseignement de ces notions.

5 Annexe : la démonstration correspondant au paragraphe 3.5 (méthode de Woolf)

Estimation asymptotique par intervalle de la cote du succès en utilisant la méthode Delta et la méthode de Wald.

Pour cela, notons $\phi : x \mapsto \ln\left(\frac{x}{1-x}\right)$.

Étant donné X qui suit la loi de Bernouilli de paramètre π, c'est-à-dire la loi binomiale $\mathfrak{B}(1,\pi)$, et $n \in \mathbb{N}$ considérons $\hat{\pi} = \frac{\sum_{i=1}^{n} X_i}{n}$, les variables indépendantes X_i ayant la même loi que X. $\hat{\pi}$ est un estimateur de π.

Déterminons une approximation normale de la loi de probabilité du logarithme néperien de la cote observée du succès (ou odds), $\phi(\hat{\pi}) = \ln\left(\frac{\hat{\pi}}{1-\hat{\pi}}\right)$. (Ceci est judicieux, car l'approximation normale du logarithme sera plus utilisable que celle de $\frac{\hat{\pi}}{1-\hat{\pi}}$, grâce à plus de similarité avec une loi normale.)

Utilisons la méthode Delta, illustrée par le schéma ci-contre, qui est extrait du livre d'Alan Agresti.

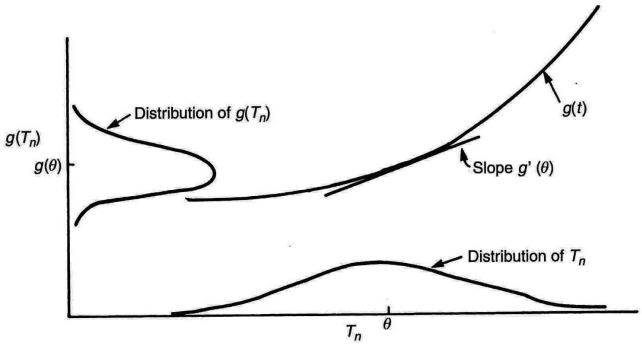


Figure 3.1 Depiction of delta method.

D'après le théorème central limite, la loi de probabilité asymptotique de $\sqrt{n}(\hat{\pi}-\pi)$ est $\mathscr{N}(0,\pi(1-\pi))$.

La méthode Delta part de l'approximation issue du développement d'ordre 1 : $\sqrt{n}(\phi(\hat{\pi})-\phi(\pi)) \cong \sqrt{n}(\hat{\pi}-\pi)\phi'(\pi)$.

La loi de probabilité asymptotique de $\sqrt{n}(\phi(\hat{\pi})-\phi(\pi))$, c'est-à-dire de $\sqrt{n}\left(\ln\left(\frac{\hat{\pi}}{1-\hat{\pi}}\right) - \ln\left(\frac{\pi}{1-\pi}\right)\right)$, est donc $\mathscr{N}\left(0,\pi(1-\pi)\times(\phi'(\pi))^2\right)$.

Une approximation normale de la loi de probabilité de $\ln\left(\frac{\hat{\pi}}{1-\hat{\pi}}\right)$ est donc $\mathscr{N}\left(\ln\left(\frac{\pi}{1-\pi}\right), \frac{\pi(1-\pi)}{n}\times(\phi'(\pi))^2\right)$

Et $\pi(1-\pi)\times(\phi'(\pi))^2 = \pi(1-\pi)\times\left(\frac{1}{\pi(1-\pi)}\right)^2 = \frac{1}{\pi(1-\pi)} = \frac{1}{\pi}+\frac{1}{1-\pi}$.

Disons maintenant que la notation $\hat{\pi}$ ne représente plus la variable aléatoire $\frac{\sum_{i=1}^{n} X_i}{n}$, mais une réalisation de cette variable aléatoire, c'est-à-dire le nombre obtenu lors d'une expérience.

D'après la normalité asymptotique ci-dessus, on a obtenu, pour $\ln\left(\frac{\pi}{1-\pi}\right)$ au niveau 95%, l'intervalle de bornes : $\ln\left(\frac{\hat{\pi}}{1-\hat{\pi}}\right) \pm 1.96\times\sqrt{\frac{1}{n\pi}+\frac{1}{n(1-\pi)}}$.

Mais la demi-largeur de cet intervalle dépend du nombre inconnu π. La méthode de Wald consiste à remplacer, dans son expression, π par $\hat{\pi}$. D'où l'intervalle de confiance pour $\ln\left(\frac{\pi}{1-\pi}\right)$:

$$\left[\ln\left(\frac{\hat{\pi}}{1-\hat{\pi}}\right) - 1.96\times\sqrt{\frac{1}{n\hat{\pi}}+\frac{1}{n(1-\hat{\pi})}}\ ;\ \ln\left(\frac{\hat{\pi}}{1-\hat{\pi}}\right) + 1.96\times\sqrt{\frac{1}{n\hat{\pi}}+\frac{1}{n(1-\hat{\pi})}}\right]$$

ou encore

$$\left[\ln\left(\frac{\hat{\pi}}{1-\hat{\pi}}\right) - 1.96\times\sqrt{\frac{1}{a}+\frac{1}{b}}\ ;\ \ln\left(\frac{\hat{\pi}}{1-\hat{\pi}}\right) + 1.96\times\sqrt{\frac{1}{a}+\frac{1}{b}}\right]$$

où a est le nombre de succès obtenus parmi les n réalisations de X et où $b = n-a$.

Enfin, on obtient un intervalle de confiance (asymétrique) de niveau 95% pour $\frac{\pi}{1-\pi}$ en prenant l'exponentielle de chaque borne, ce qui donne :

$$\left[\frac{a}{b}\times e^{-1.96\times\sqrt{\frac{1}{a}+\frac{1}{b}}}\ ;\ \frac{a}{b}\times e^{1.96\times\sqrt{\frac{1}{a}+\frac{1}{b}}}\right]$$

Estimation asymptotique par intervalle du rapport de deux cotes en utilisant les mêmes méthodes.

Il s'agit d'obtenir un intervalle de confiance asymptotique pour $\frac{\pi_1(1-\pi_2)}{\pi_2(1-\pi_1)}$; considérons d'abord $\ln\left(\frac{\hat{\pi}_1(1-\hat{\pi}_2)}{\hat{\pi}_2(1-\hat{\pi}_1)}\right) = \ln\left(\frac{\hat{\pi}_1}{1-\hat{\pi}_1}\right) - \ln\left(\frac{\hat{\pi}_2}{1-\hat{\pi}_2}\right)$.

D'après ce qui précède, les variables $\hat{\pi}_1$ et $\hat{\pi}_2$ étant indépendantes, une approximation normale de la loi de probabilité de $\ln\left(\frac{\hat{\pi}_1(1-\hat{\pi}_2)}{\hat{\pi}_2(1-\hat{\pi}_1)}\right)$ est $\mathscr{N}\left(\phi(\pi_1)-\phi(\pi_2), \mathbb{V}(\hat{\pi}_1)\times(\phi'(\pi_1))^2 + \mathbb{V}(\hat{\pi}_2)\times(\phi'(\pi_2))^2\right)$.

En répétant ce qui précède pour $\hat{\pi}_1$ et pour $\hat{\pi}_2$, on obtient un intervalle de confiance asymptotique de niveau 95% pour le rapport de cotes :

$$\left[\frac{ad}{bc}\times e^{-1.96\times\sqrt{\frac{1}{a}+\frac{1}{b}+\frac{1}{c}+\frac{1}{d}}}\ ;\ \frac{ad}{bc}\times e^{1.96\times\sqrt{\frac{1}{a}+\frac{1}{b}\,\frac{1}{c}+\frac{1}{d}+}}\right]$$

6 Bibliographie

Bibliographie

1. André Dufour : *Le tabagisme en France*. Librairie Séguier, 1990
2. Bulletin épidémiologique hebdomadaire, 28 mai 2013, num. 20-21,
   http ://www.invs.sante.fr/content/download/65767/255731/version/10/file/BEH_20-21.pdf
3. Catherine Hill : *Mortalité attribuable au tabac en France*, 2015,
   http ://www.sante.gouv.fr/IMG/pdf/Mortalite_attribuable_au_tabac_en_France.pdf
4. Bulletin épidémiologique hebdomadaire, 29 mai 2015, num. 17-18,
   http ://www.invs.sante.fr/beh/2015/17-18/pdf/2015_17-18.pdf
5. Didier Nourrisson : *Histoire sociale du tabac*. Collection Vivre l'histoire, Christian éditeur, 1999
6. Alan Agresti : Categorical Data Analysis, Wiley, 2013
7. Marie-Claude VIANO et Charles SUQUET : Éléments de Statistique Asymptotique,
   http ://math.univ-lille1.fr/ suquet/Polys/StAs10.pdf
8. Robert N. Proctor : *La guerre des nazis contre le cancer*. Les Belles Lettres, 2001
9. Robert N. Proctor : *Golden Holocaust, la conspiration des indistriels du tabac*. Édition des Équateurs, 2014
10. Karl Pearson. *On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling*. Philosophical Magazine, 1900.
11. G. Udny Yule. *On the Association of Attributes in Statistics* Philosophical Transactions of the Royal Society of London. Series A, Vol. 194 (1900), pp. 257-319, https ://archive.org/download/philtrans08815622/08815622.pdf.
12. Ronald A. Fisher. *Statistical Methods for Research Workers*. Oliver and Boyd, Edinburgh, 1925,
    http ://www.haghish.com/resources/materials/Statistical_Methods_for_Research_Workers.pdf.
13. Austin B. Hill. *Principles of Medical Statistics*. Oxford University Press, 1937.
14. Franz H. Müller. *Tabakmissbrauch und Lungencarcinom*. Zeitschrift fur Krebsforschung, 1939. 49 ; p. 57-85.
15. Ernst L. Wynder and Evarts A. Graham. *Tobacco Smoking as a Possible Factor in Bronchiogenic Carcinoma*. Journal of the American Medical Association, 1950,
    http ://www.epidemiology.ch/history/PDF%20bg/
    Wynder%20and%20Graham%201950%20tobacco%20smoking%20as%20a%20possible%20etiologic.pdf
16. Richard Doll and Austin B. Hill. *Smoking and Carcinoma of the Lung*. British Medical Journal, 1950,
    http ://www.ncbi.nlm.nih.gov/pmc/articles/PMC2038856/pdf/brmedj03566-0003.pdf
17. Jerome Cornfield. *A method of estimating comparative rates from clinical data*. Applications to cancer of the lung, breast, and cervix. JOURNAL OF THE NATIONAL CANCER INSTITUTE, 1951,
    http ://arch.neicon.ru/xmlui/bitstream/handle/123456789/4121050/JNCIjnci_11_6_11-6-1269.pdf ?sequence=1
18. J. Berkson. *Smoking and Lung Cancer : Some Observations on two Recent Reports*. American Statistical Association Journal, 1958. p. 28-38.
19. Cristian Tomasetti and Bert Vogelstein, *Variation in cancer risk among tissues can be explained by the number of stem cell divisions* ,Science, 2 janvier 2015, vol 347, issue 6217

et quelques URL nous concernant :

http ://www.math.unicaen.fr/irem/spip.php ?rubrique43
http ://jacques.faisant.pagesperso-orange.fr/JN2014/
http ://www.math.unicaen.fr/irem/spip.php ?article6
http ://jacques.faisant.pagesperso-orange.fr/JN2015/