

I. Les calculs rénaux	1
I.1 C'est quoi ?	1
I.2 Une étude sur des traitements contre les calculs rénaux	2
II. D'autres exemples	5
II.1 Fumer est bon pour la santé	5
II.2 Discrimination sexuelle dans une université ?	6
II.3 Trois exemples divers : notes d'élèves, taux de mortalité et accidents de la route	6
II.4 Des condamnations racistes aux U.S.A. dans les années 1970	7
II.5 Statistiques sportives : qui est le meilleur joueur ?	8
II.6 Statistiques politiques et citoyennes : le Civil Right Act	8
III. Quand le paradoxe se produit-il ?	9
III.1 Explication mathématique	9
III.1 Explication statistique	9
IV. Comment se prémunir du paradoxe de Simpson ?	10
V. D'autres paradoxes statistiques	11
V.1 Salaires qui baissent mais qui augmentent... Syndicats VS patronat	11
V.2 Sondage : le paradoxe du nombre moyen d'enfants	11

I. Les calculs rénaux

I.1 C'est quoi ?

Les calculs rénaux, communément appelés « pierres aux reins », sont des **cristaux durs qui se forment dans les reins** et peuvent entraîner de vives douleurs. Les médecins utilisent le terme *lithiase urinaire* pour désigner ces cristaux, qui peuvent aussi se retrouver dans le reste de l'appareil urinaire : dans la vessie, l'urètre ou les uretères.

Dans près de 90 % des cas, les calculs urinaires se forment à l'intérieur d'un rein. La plupart d'entre eux (80 %) s'éliminent spontanément en passant à travers les divers conduits du système urinaire et n'entraînent que peu de symptômes.

Cependant, les uretères, situés entre les reins et la vessie, sont des conduits de très petit calibre. Un calcul formé dans le rein, qui est en transit vers la vessie, peut facilement obstruer un uretère et ainsi causer de vives douleurs.

Les calculs rénaux sont très fréquents, et leur prévalence semble avoir augmenté au cours des 30 dernières années. De 5 % à 10 % des individus vivront une crise de colique néphrétique au cours de leur vie. Les calculs rénaux surviennent plus fréquemment dans la quarantaine. Ils sont **2 fois plus fréquents chez l'homme que chez la femme**.

Les calculs sont le résultat de la **crystallisation de sels minéraux et d'acides** présents en trop forte concentration dans l'urine.

Le plus souvent, ils sont dus à un manque de dilution des urines, c'est-à-dire à une consommation trop faible d'eau.

Une alimentation déséquilibrée, trop riche en sucre ou en protéines, peut également être en cause.

Plus rarement, une infection, certains médicaments, une maladie génétique ou métabolique (diabète) peuvent entraîner la formation de calculs urinaires.

Source : http://www.passeportsante.net/fr/Maux/Problemes/Fiche.aspx?doc=lithiase_urinaire_pm

I.2 Une étude sur des traitements contre les calculs rénaux

En 1986, C. Charig, D. Webb, S. Payne et J. Wickham publient dans la revue *British Medical Journal* un article titré *Comparison of treatment of renal calculi by open surgery, percutaneous nephrolithotomy, and extracorporeal shockwave lithotripsy*¹, c'est-à-dire *Comparaison du traitement des calculs rénaux par chirurgie ouverte*², *néphrolithotomie percutanée*³ et *lithotripsie extracorporelle par ondes de choc*⁴.

En voici un premier extrait :

This study was designed to compare different methods of treating renal calculi in order to establish which was the most cost effective and successful. Of 1052 patients with renal calculi, 350 underwent open surgery, 350 percutaneous nephrolithotomy, 328 extracorporeal shockwave lithotripsy (ESWL), and 24 both percutaneous nephrolithotomy and ESWL. Treatment was defined as successful if stones were eliminated or reduced to less than 2 mm after three months. Success was achieved in 273 (78 %) patients after open surgery, 289 (83 %) after percutaneous nephrolithotomy, 301 (92 %) after ESWL, and 15 (62 %) after percutaneous nephrolithotomy and ESWL. Comparative total costs to the NHS were estimated as 3500 pounds for open surgery, 1861 pounds for percutaneous nephrolithotomy, 1789 pounds for ESWL, and 3210 pounds for both ESWL and nephrolithotomy. ESWL caused no blood loss and little morbidity and is the cheapest and quickest way of returning patients to normal life.

1. Quel est l'objectif principal de l'étude ?
2. Après la lecture du premier extrait, quel traitement vous semble être le meilleur ?

Un deuxième extrait :

With any new and capittally expensive treatment three questions need to be answered : what is the difference in mortality ? What is the difference in morbidity ? Which mode of treatment is the most cost effective ? We tried to define the current place of extracorporeal shockwave lithotripsy (ESWL) in the management of renal calculi by comparing 350 cases of open stone removal, 350 cases of percutaneous nephrolithotomy, and 352 cases of ESWL. All patients were treated by the same team of surgeons under the direct supervision of one consultant. Patients in all groups were treated consecutively over 14 years (open surgery from 1972 to 1980, percutaneous nephrolithotomy from 1980 to 1985, and ESWL in 1985).

1 <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1339981/pdf/bmjcred00227-0031.pdf>

2 Chirurgie « classique » : en ouvrant le ventre.

3 Le calcul rénal est situé dans les cavités du rein. La néphrolithotomie percutanée consiste à introduire directement dans le rein à travers la peau et la paroi du dos un appareil qui permet de voir, de fragmenter et d'enlever le ou les calculs.

Cette intervention nécessite habituellement une anesthésie générale et une hospitalisation de 4 à 5 jours au total.

4 Méthode qui permet de fragmenter le calcul grâce à des ondes de chocs. Les ondes de choc sont générées par un lithotripteur au sein d'une tête de traitement. Le contact de la tête de traitement avec la peau du patient permet de propager et de focaliser les ondes de chocs, à l'intérieur du corps humain, jusqu'au calcul.

Le calcul est ainsi réduit en un très grand nombre de fragments qui sont ensuite éliminés par les voies naturelles.

3. Quelle est la motivation principale des auteurs de l'étude ?
4. Cette étude vous semble-t-elle sérieuse ? Argumenter.

Concentrons-nous sur la chirurgie ouverte (traitement A) et la néphrolithotomie percutanée (traitement B). On peut lire dans l'étude :

For the purpose of this study all details of patients were assessed three months after treatment. The procedures were defined as successful if patients were free of stones at three months or, in the case of ESWL, if stones had been fragmented into particles of less than 2 mm. For each type of treatment the patients were considered in two groups : group 1 were those with stones of a mean diameter of less than 2 cm and group 2 those with stones of a mean diameter of 2 cm or more or with multiple stones.

Dans l'étude, les auteurs donnent les résultats suivants :

TABLE I—Details of patients in each treatment group

	No (%) of patients with stones		Total	Mean age (range)	M:F (%)
	<2 cm (group 1)	≥2 cm (group 2)			
Nephrolithotomy/pyelolithotomy	13 (6)	218 (94)	231	45 (12-78)	45:55
Pyelolithotomy	31 (41)	45 (59)	76	47 (16-72)	51:49
Ureterolithotomy	43 (100)		43	46 (20-68)	69:31
Percutaneous nephrolithotomy	270 (77)	80 (23)	350	52 (23-72)	68:32
ESWL	204 (62)	124 (38)	328	48 (22-83)	70:30
Percutaneous nephrolithotomy and ESWL		24 (100)	24		

TABLE II—Success rate of treatment* (figures are numbers (%) of patients)

	Group 1	Group 2	Overall
Nephrolithotomy/pyelolithotomy	12 (92)	154 (71)	166 (72)
Pyelolithotomy	26 (84)	38 (84)	64 (84)
Ureterolithotomy	43 (100)		43 (100)
All open procedures	81 (93)	192 (73)	273 (78)
Percutaneous nephrolithotomy†	234 (87)	55 (69)	289 (83)
ESWL	200 (98)	101 (82)	301 (92)
Percutaneous nephrolithotomy and ESWL		15 (62)	15 (62)

*Success defined as no stones at three months or stone reduced to particles <2 mm in size.

†52 with electrohydraulic lithotripsy, 69 with ultrasound.

Après quelques calculs, on obtient donc le tableau suivant :

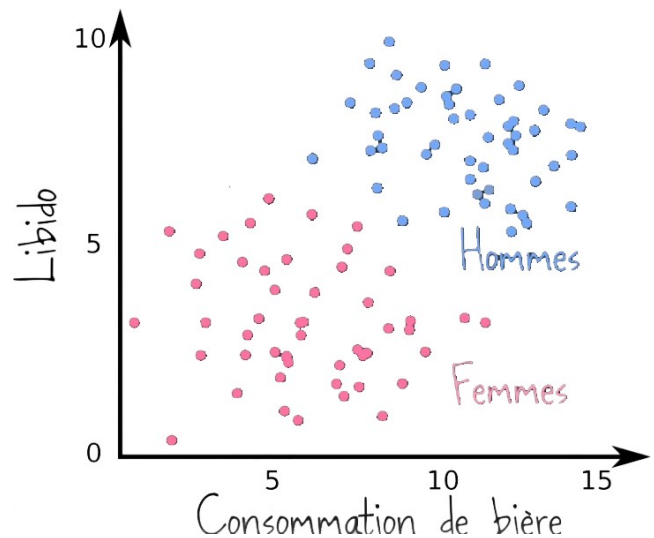
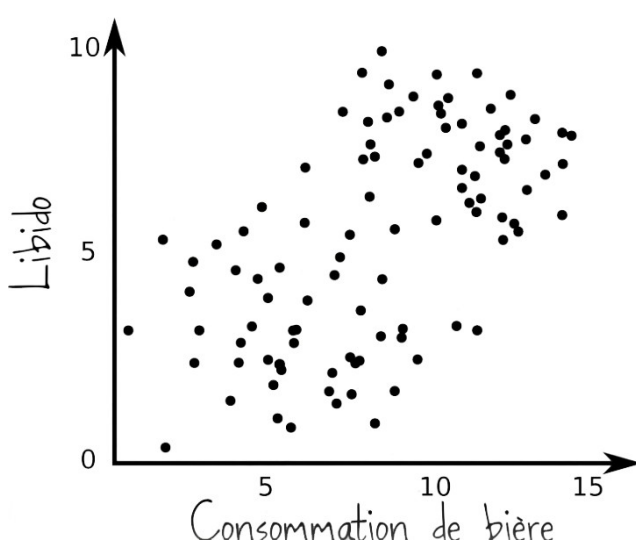
	Traitement A	Traitement B
Petits calculs (<2cm)	81/87 93%	234/270 87%
Gros calculs (≥2cm)	192/263 73%	55/80 69%
Total		

5. a) Si un patient a des petits calculs, quel traitement fonctionne le mieux ?
- b) Et si le patient a de gros calculs ?
- c) Finalement, quel traitement semble le meilleur ?
6. Compléter la ligne *Total*. Quel traitement semble le meilleur ?

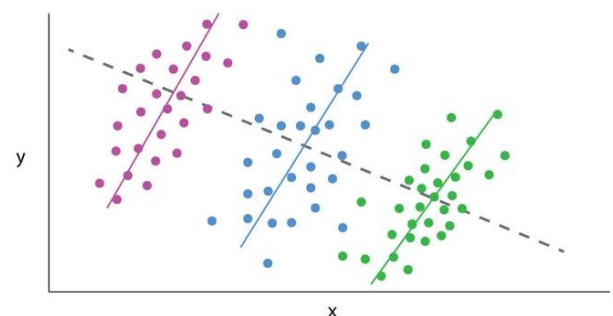
Il n'y a aucune entourloupe statistique ou aucune manipulation, ce que vous lisez là, c'est bien la réalité des chiffres. Vous avez là un bel exemple de ce qu'on appelle un **paradoxe de Simpson**.

- Le paradoxe de Simpson ou effet de Yule-Simpson est un paradoxe statistique décrit par Edward H. Simpson en 1951, mais les statisticiens Karl Pearson (1899) et Udny Yule (1903) avaient mentionné des faits similaires dans lesquels un phénomène observé de plusieurs groupes semble s'inverser lorsque les groupes sont combinés.
- Ce résultat qui semble impossible au premier abord est lié à des éléments qui ne sont pas pris en compte (comme la présence de variables non-indépendantes ou de différences d'effectifs entre les groupes, etc.). Il est souvent rencontré dans la réalité, en particulier dans les sciences sociales et les statistiques médicales.
- Le terme *paradoxe de Simpson* a été introduit en 1972 par Colin R. Blyth.

Je vous encourage à regarder cette vidéo de David Louapre (alias ScienceEtonnante sur YouTube) :
https://youtu.be/vs_Zzf_vL2I



Finalement, ce paradoxe se résume en une seule image :



Source de l'image : <https://economicthinkinginaction.com/>

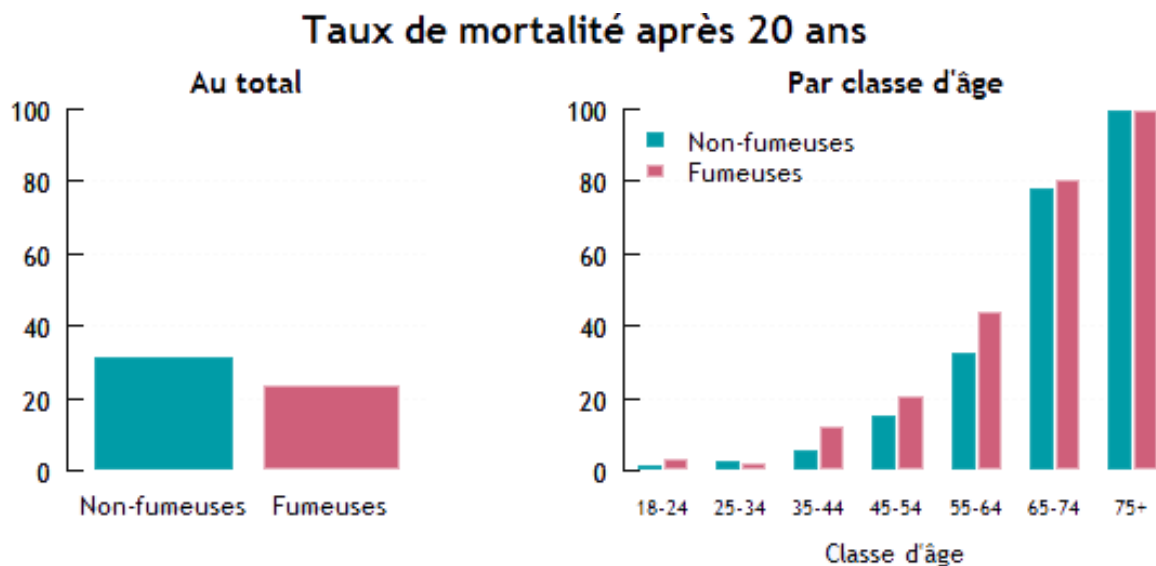
II. D'autres exemples

II.1 Fumer est bon pour la santé

Dans les années 1970, les femmes d'un village de Wickham (ville près de Newcastle en Angleterre) ont participé à une étude⁵ : 1314 femmes ont été suivies pendant 20 ans, et l'objectif était de comparer le taux de mortalité des fumeuses et des non-fumeuses.

Après 20 ans, le taux de mortalité chez les fumeuses était de 24 %, celui des non-fumeuses était 31 %. Alors, est-ce que ne pas fumer tue ?!

En réalité, il y avait 582 fumeuses et 139 sont mortes (soit 24 %), ainsi que 732 non-fumeuses dont 230 sont mortes (soit 31 %) ... Mais en séparant par classe d'âge :



Dans chaque tranche la mortalité chez les fumeuses a été supérieure à celle des non-fumeuses. On est rassurés, mais comment les chiffres peuvent-ils s'inverser quand on groupe tout le monde ?

Dans la population initiale, il y avait plus de femmes âgées chez les non-fumeuses que chez les fumeuses. Même si dans chaque tranche d'âge les non-fumeuses meurent moins, cet effet est compensé par le fait que la tranche d'âge « élevée » est sur-représentée chez les non-fumeuses, qui donc en moyenne meurent plus !

Source : <https://sciencetonnante.wordpress.com/2013/04/29/le-paradoxe-de-simpson/>

⁵ D. Appleton, J. French et M. Vanderpump. « Ignoring a covariate : An example of Simpson's paradox. » The American Statistician 50.4 (1996) : 340-341.

II.2 Discrimination sexuelle dans une université ?

En 1973, l'université de Berkeley fut poursuivie pour discrimination envers les filles. L'affaire semblait claire. Parmi les candidates, seules 35 % étaient retenues alors que 44 % des candidatures masculines l'étaient. L'étude a été précisée sur les six départements les plus importants, notés ici de A à F.

Département	Garçons	Admis	Filles	Admises
A	825	62 %	108	82 %
B	560	63 %	25	68 %
C	325	37 %	593	34 %
D	417	33 %	375	35 %
E	191	28 %	393	24 %
F	272	6 %	341	7 %

Détails des admissions.

Ce tableau ne montre aucune discrimination envers les femmes. Au contraire, le taux d'admission des filles dans le principal département (A) est nettement supérieur à celui des garçons. L'explication vient quand on regarde le nombre de candidatures dans ces départements. Les femmes semblent avoir tendance à postuler en masse à des départements très sélectifs. Dans ceux-ci, leur taux d'admission est à peine plus faible que celui des hommes. Dans les autres, elles sont plus largement sélectionnées que les hommes. Quand on fait la moyenne globale, ce sont les départements sélectifs qui ont plus de poids, puisqu'elles y postulent en masse. Ce paradoxe a été étudié par Edward Simpson (né en 1922). Il n'est pas rare, on le retrouve même souvent.

Source : Tangente Hors-Série n°49, Les maths de l'impossible

II.3 Trois exemples divers : notes d'élèves, taux de mortalité et accidents de la route

Pris parmi des dizaines d'autres, voici trois cas réels où le paradoxe de Simpson a produit des surprises et a contraint à analyser plus finement ce que des manipulations statistiques imprudentes faisaient apparaître.

Aptitudes en progrès pour tous, mais dont la moyenne est stable.

Aux États-Unis, au cours des années 1981-2002, la moyenne des notes obtenues par les élèves soumis au test linguistique SAT (*Scholastic Aptitude Test*) a été stable, avec un résultat toujours proche de 504 points. Pourtant, durant cette période, la moyenne pour le test mesuré séparément par groupe ethnique s'est accrue dans chaque groupe d'au moins huit points.

Gerald Bracey, de l'Université George Mason, a étudié cet étrange phénomène et a conclu qu'on était

là en présence d'un cas du paradoxe de Simpson.

La clef de l'explication se trouve dans le fait que pendant la période considérée, le nombre d'élèves blancs soumis au test a proportionnellement décré, et que la proportion d'élèves appartenant à des minorités (réussissant en général moins bien le test linguistique) a augmenté, conduisant au total à une stagnation des résultats globaux qui, pourtant, progressaient dans chaque catégorie.

Mauvais taux de mortalité dans chaque tranche d'âge, mais globalement bons.

En 1986, Joel Cohen, de l'Université Rockefeller à New York, a identifié un cas réel de paradoxe de Simpson dans le domaine de la démographie. Il s'est intéressé à la mortalité au Costa Rica et en Suède (connue

pour son excellente espérance de vie). Sans surprise, il a trouvé qu'en 1960, le taux de mortalité des femmes dans toute tranche d'âge fixée était supérieur au Costa Rica, comparé au taux équivalent en Suède. Pourtant, le taux de mortalité général des femmes au Costa Rica était inférieur à celui de la Suède.

L'explication est encore liée au paradoxe de Simpson, et elle provient de la structure différente des populations. La population du Costa Rica est beaucoup plus jeune en moyenne que celle de la Suède, et donc les jeunes classes d'âge (qui ont un faible taux de mortalité) pèsent plus dans la moyenne pour le Costa Rica que pour la Suède, conduisant à un taux de mortalité global assez faible au Costa Rica, malgré un taux assez mauvais dans chaque tranche d'âge.

En roulant plus vite, vous aurez moins d'accidents.

Gary Davis, de l'Université du Minnesota, étudiait en 2004 la relation entre le nombre d'accidents piéton-véhicule et la vitesse moyennée des véhicules en divers endroits d'une ville.

Il proposa un modèle de cette relation qu'il voulait utiliser pour évaluer l'intérêt de limiter plus sévèrement la vitesse autorisée des véhicules. De façon inattendue, le modèle montrait qu'en faisant passer la limite de 30 miles par heure à 25, le nombre d'accidents augmenterait.

Là encore, bien que pour chaque type de sites, le nombre d'accidents diminuait quand la vitesse limite était réduite, une agrégation malheureuse des données (qui ne prenait pas en compte que le nombre d'accidents était bien plus rare dans les zones résidentielles) engendrait une conclusion absurde.



Étudiants en Virginie



Rue au Costa Rica



Rue en Suède



Limitation de vitesse

Source : J.-P. Delahaye, Pour la Science - n° 429 - Juillet 2013

II.4 Des condamnations racistes aux U.S.A. dans les années 1970

Le 11 mars 1979, le New York Times publie les données concernant 4764 homicides perpétrés en Floride entre 1973 et 1979. Les variables considérées sont la couleur de peau du meurtrier (blanc ou noir) et la condamnation à la fin du jugement (peine de mort ou non). Les chiffres en question étaient les suivants :

Meurtrier	Peine de mort	Autre peine
Blanc	72	2185
Noir	59	2448

L'idée était de savoir qui des blancs ou des noirs étaient le plus souvent condamnés à mort et, plus généralement, s'il y avait un problème de racisme dans la justice américaine.

Un calcul rapide amène à **3,19 % des meurtriers blancs** qui sont condamnés à mort, contre **2,35 % des meurtriers noirs** qui sont condamnés à mort.

Les meurtriers blancs sont donc aussi sévèrement punis que les meurtriers noirs...

La justice américaine n'était donc pas raciste ?

Quand on a une vague idée du contexte aux États-Unis dans les années 1970, on peut en douter !

Et si on reprenait le tableau initial en distinguant la couleur de peau de la victime :

Meurtrier	Victime	Peine de mort	Autre peine
Blanc	Blanche	72	2074
Blanc	Noire	0	111
Noir	Blanche	48	239
Noir	Noire	11	2209

Si la victime est blanche :

3,36 % des meurtriers blancs sont condamnés à mort ($72 \div 2146$).

16,72 % des meurtriers noirs sont condamnés à mort ($48 \div 287$).

Donc les meurtriers noirs sont beaucoup plus sévèrement punis que les meurtriers blancs.

Si la victime est noire :

0 % des meurtriers blancs sont condamnés à mort ($0 \div 111$).

0,48 % des meurtriers noirs sont condamnés à mort ($11 \div 2310$).

Donc les meurtriers noirs sont légèrement plus sévèrement punis que les meurtriers blancs.

... en cachant une variable, l'interprétation des chiffres peut mener bien loin de la réalité.

Source : <http://www.cahiersdufootball.net/article-limites-et-distorsions-statistiques-5841>

II.5 Statistiques sportives : qui est le meilleur joueur ?

Un autre exemple célèbre est donné par les statistiques de réussite au baseball.

En 1995, le taux de succès du batteur David Justice a été meilleur que celui de son adversaire Derek Jeter (25,3 % contre 25,0 %).

L'année suivante, en 1996, même résultat : Justice surpasse Jeter par 32,1 % contre 31,4 %.

Et pourtant, si on combine l'ensemble des deux années 1995 et 1996 :

Derek Jeter a été le meilleur avec 31 % contre 27 % !

II.6 Statistiques politiques et citoyennes : le Civil Right Act

En 1964 les États-Unis ont voté une loi historique, le Civil Right Act, qui fut un pas déterminant vers l'abolition de la ségrégation raciale.

Sur l'ensemble du pays, 80 % des républicains ont voté en sa faveur, contre 61 % des démocrates.

Étonnant quand on connaît les positions de ces deux partis.

Et pourtant si on distribue ces résultats entre les états du Nord et du Sud du pays, aussi bien au Nord qu'au Sud les démocrates ont plus voté que les républicains en faveur de la loi !

III. Quand le paradoxe se produit-il ?

III.1 Explication mathématique

Voici une configuration générale :

Peut-on avoir $\frac{a}{c} < \frac{A}{C}$ et $\frac{b}{d} < \frac{B}{D}$ mais $\frac{a+b}{c+d} > \frac{A+C}{B+D}$?

Bien sûr que oui...

Aucune règle de calcul permet de conclure que $\frac{a+b}{c+d} < \frac{A+C}{B+D}$.

	Caractère A	Caractère B
Sous-groupe 1	$\frac{a}{c}$	$\frac{A}{C}$
Sous-groupe 2	$\frac{b}{d}$	$\frac{B}{D}$
<u>Total</u>	$\frac{a+b}{c+d}$	$\frac{A+C}{B+D}$

On peut donc trouver des contre-exemples comme ceux examinés dans ce devoir.

D'après J.-P. Delahaye, la proportion de jeux de huit entiers pris entre 1 et n vérifiant les inégalités du paradoxe tend vers 0,9606 % quand n tend vers l'infini : même si les cas où le paradoxe de Simpson se produit sont gênants, ils sont heureusement assez rares.

III.1 Explication statistique

Pour que le paradoxe se produise, il faut deux ingrédients :

- il faut une variable qui influe sur le résultat final (le « groupe »), et qui n'est pas forcément explicitée au départ. On appelle cela un ***facteur de confusion***⁶.

Il s'agit de la taille des calculs dans le premier exemple, car celle-ci influe sur la probabilité de succès du traitement, et de l'âge des personnes dans le second exemple, lequel évidemment joue sur la mortalité.

- il faut que l'échantillon qu'on étudie ***ne soit pas distribué de manière homogène*** : dans le cas du tabac, il y a plus de vieilles femmes dans l'échantillon des non-fumeuses que chez les fumeuses ; dans le cas des calculs rénaux, le traitement A est plus souvent donné sur les gros calculs et le « B » sur les petits.

Quand ces deux conditions sont réunies, le paradoxe de Simpson peut se produire !

À cause de la distribution hétérogène de l'échantillon, regrouper les données pointe une tendance qui peut être fausse, et qui disparaît si on analyse les données en séparant selon le facteur de confusion.

6 Pour bien comprendre ce qu'est un facteur de confusion, voir cette vidéo de David Louapre (*Science étonnante*) avec un exemple très visuel de bière qui influencerait sur la libido : https://youtu.be/vs_Zzf_vL2I

IV. Comment se prémunir du paradoxe de Simpson ?

J'imagine que vous voyez aisément *le potentiel de manipulation qui se cache derrière ce paradoxe* : on peut vous faire croire à quelque chose (le chômage a baissé, tel traitement marche mieux, tel individu est meilleur, etc.) alors qu'en regardant les chiffres dans le détail, les effets peuvent disparaître ou s'inverser ! Alors que faire ?

Tout d'abord, il faut se rappeler : cet effet se produit *quand il existe une variable cachée influente*, et que *l'échantillon sur lequel on se base n'est pas homogène*. En sciences, c'est pour cela que l'on préfère en général des expériences « randomisées », qui permettent d'assurer une distribution homogène : par exemple si vous avez des calculs rénaux et que vous participez à une expérience pour comparer les traitements, on vous assigne au hasard le traitement A ou B, sans que la taille des calculs influe sur la décision. On gomme ainsi l'inhomogénéité de distribution, et le paradoxe disparaît : le traitement A sera bien vu comme étant le meilleur.

Quand on vous présente des chiffres, il faut donc avoir l'œil critique, et être particulièrement méfiants quand ces chiffres sont issus de données analysées *a posteriori*, plutôt que sur un échantillon expérimental qu'on a soi-même construit *a priori* (en randomisant).

Réfléchissez au point suivant : conclure que « le lit est l'endroit le plus dangereux du monde, c'est là que la plupart des gens meurent » c'est se tromper, car on utilise des données non-randomisées.

Enfin rappelez-vous, ce paradoxe se produit quand il existe une variable cachée fortement influente. Cela signifie que les chiffres bruts ont peu de sens, et doivent être critiqués par un expert du domaine, susceptible de pointer l'existence d'un tel facteur. A l'heure où fleurit la mode du « fact-checking », on a un peu tendance à nous faire croire que les chiffres seraient la vérité « nue »... Non, la vérité nue n'existe pas, et on aura toujours besoin de gens au courant pour interpréter correctement des chiffres, qu'ils soient scientifiques, économiques ou médicaux.

Source : <https://sciencetonnante.wordpress.com/2013/04/29/le-paradoxe-de-simpson/>

V. D'autres paradoxes statistiques

V.1 Salaires qui baissent mais qui augmentent... Syndicats VS patronat

Au sein d'une entreprise, un conflit oppose les syndicats et le patron.

Les syndicats : « les salaires des ouvriers et ceux des cadres ont baissé cette année de 10 % ».

Le patron : « nos calculs indiquent que le salaire moyen dans l'entreprise a augmenté de 152 % ! »

		Ouvriers	Cadres
2006	Salaire	200 €	2000 €
	Effectif	1000	100
2007	Salaire	180 €	1800 €
	Effectif	600	500

Pourtant, personne ne ment. Explications ? Faites les calculs...

L'arnaque est que les effectifs par catégorie ont changé entre 2006 et 2007 : les ouvriers sont désormais moins nombreux et les cadres plus nombreux. Il en résulte que la baisse du salaire dans chaque catégorie est compensée par l'augmentation du nombre des cadres qui sont mieux rétribués.

Une baisse de 10 % du salaire de chaque catégorie de personnel est parfaitement compatible avec une augmentation du salaire moyen des employés.

V.2 Sondage : le paradoxe du nombre moyen d'enfants

Une enquête exhaustive menée dans une ville lointaine indique que les familles ayant des enfants de moins de 18 ans se répartissent de la manière suivante : 10 % de familles à 1 enfant, 50 % de familles à 2 enfants, 30 % de familles à 3 enfants, 10 % de familles à 4 enfants.

Le nombre moyen d'enfants par famille (parmi celles qui ont des enfants) est donc de : $\frac{10+100+90+40}{100} = 2,4$.

Pour contrôler cette statistique sur les familles ayant des enfants, les autorités administratives procèdent à un sondage : on interroge au hasard 1 000 enfants de moins de 18 ans et on leur demande combien il y a d'enfants dans leur famille⁷.

En faisant la moyenne des réponses obtenues au sondage, on obtient 2,165. Cela semble absurde.

Pourtant, en recommençant et en interrogeant cette fois 10 000 enfants, on trouve maintenant 2,164.

Un troisième sondage sur 100 000 enfants donne 2,165 à nouveau.

Pourquoi cet écart si important avec 2,4 ?

En interrogeant des enfants au hasard, vous interrogerez 4 fois plus d'enfants des familles à 4 enfants que vous n'en interrogerez dans les familles à 1 enfant, ce qui fausse la moyenne. S'il y a 1 000 familles, il y aura 100 enfants uniques, 1 000 enfants appartenant à une famille de 2 enfants, 900 enfants appartenant à une famille de 3 enfants, 400 enfants appartenant à une famille de 4 enfants. En moyenne :

$$\frac{100+1\,000\times 2+900\times 3+400\times 4}{2\,400} = 2,166.$$

Ce qu'évaluent les sondages opérés, ce n'est pas le nombre moyen d'enfants d'une famille prise au hasard, mais le nombre moyen d'enfants qu'on trouve dans la famille d'un enfant pris au hasard.

« Prendre une famille au hasard » et « Prendre un enfant au hasard » n'est pas la même chose.

Source : J.-P. Delahaye, POUR LA SCIENCE - N° 357 JUILLET 2007

⁷ Pour simplifier le problème, on imagine que dans cette ville idéale, il n'y a ni divorce, ni orphelin...