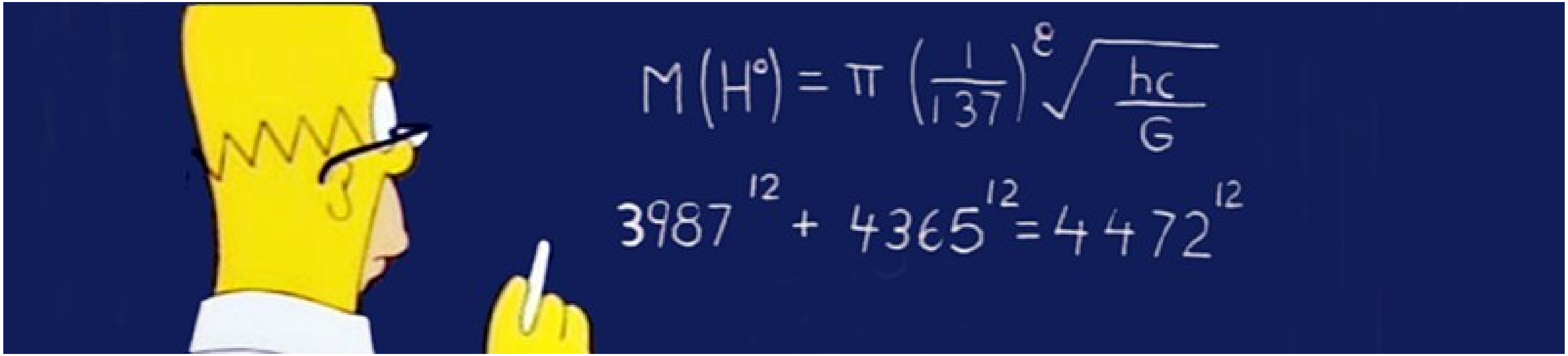
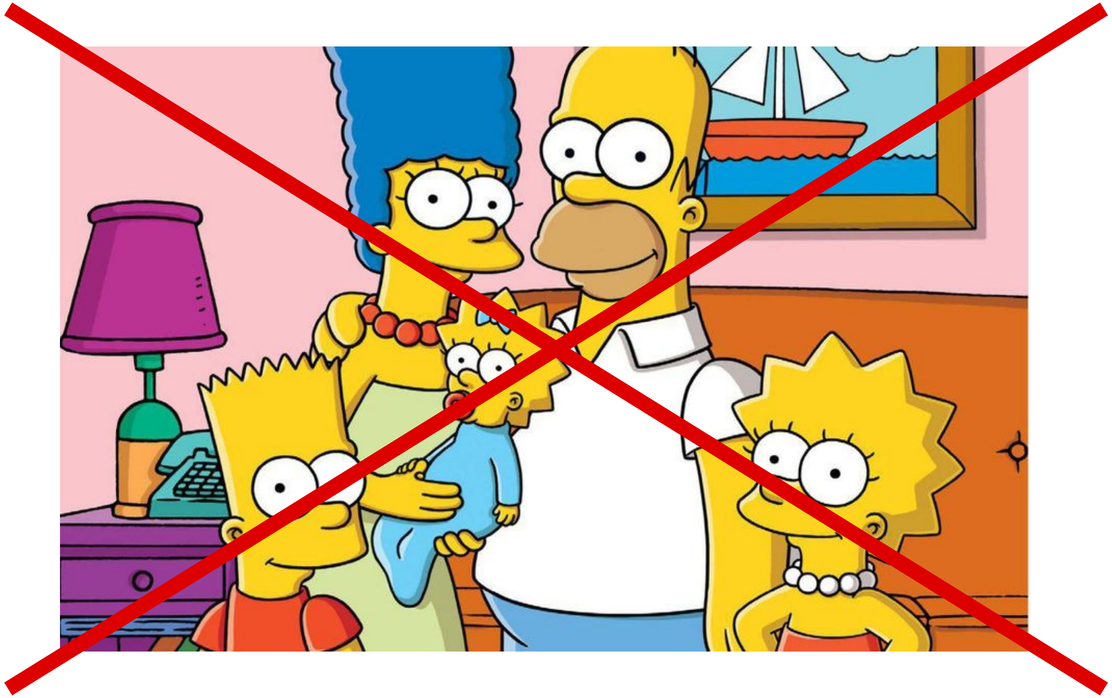
Le paradoxe de Simpson



Contrairement à ce que l’on pourrait penser, le *paradoxe de Simpson* n’est pas lié à la célèbre série « Les Simpson », mais à son fondateur Edward Simpson qui l’a expliqué pour la première fois en 1951. Ici, nous vous expliquerons donc ce phénomène.



Le paradoxe de Simpson, c’est quoi ?

Le paradoxe de Simpson est basé sur des données de statistiques mais également sur des variations de calculs qui créer des confusions. Cela aura un impact sur les choix que l’on va faire et sur la performance du résultat obtenu.

Pour mieux comprendre ce paradoxe, nous allons l’illustrer avec des exemples.

I- Calculs rénaux : quel traitement choisir ?

II- Fumer c’est bon pour la santé !

III- Lisa et Bart

Calculs rénaux : quel traitement choisir ?

Vous découvrez des calculs aux reins.Le médecin chargé de s’occuper de vous vous propose des traitements. Appelons-les traitement A et B. Le A sera une chirurgie ouverte et le B est une chirurgie qui se fait par de petits trous percés à travers la peau. Vous demandez au médecin quel traitement a le plus de taux de réussite. Il vous exposera les statistiques de succès de ces deux traitements.

Le traitement A a fonctionné dans 273 cas et le traitement B dans 289, sachant que chaque traitements a été testé sur 350 patients.

Logiquement, vous choisissez le traitement B puisqu’il a 83% de réussite, contre 79% seulement pour le traitement A.

Cependant, vous obtenez l’avis d’un autre médecin qui vous explique par un tableau que la réussite dépend de la taille des calculs.

	Traitement A	Traitement B
Petits calculs (<2cm)	81/87 93%	234/270 87%
Gros calculs (>2cm)	192/263 73%	55/80 69%
Total	273/350 78%	289/350 83%

Comme vous pouvez le constatez, si vous avez des gros calculs, le traitement A fonctionne mieux, et si vous avez des petits calculs, le traitement A est aussi le plus efficace. Cela est contradictoire avec ce que le premier médecin a dit... Et pourtant, vous avez beau compter et recompter, sur la ligne « Total », il s'agit bien des mêmes chiffres que ceux présentés par le premier médecin…

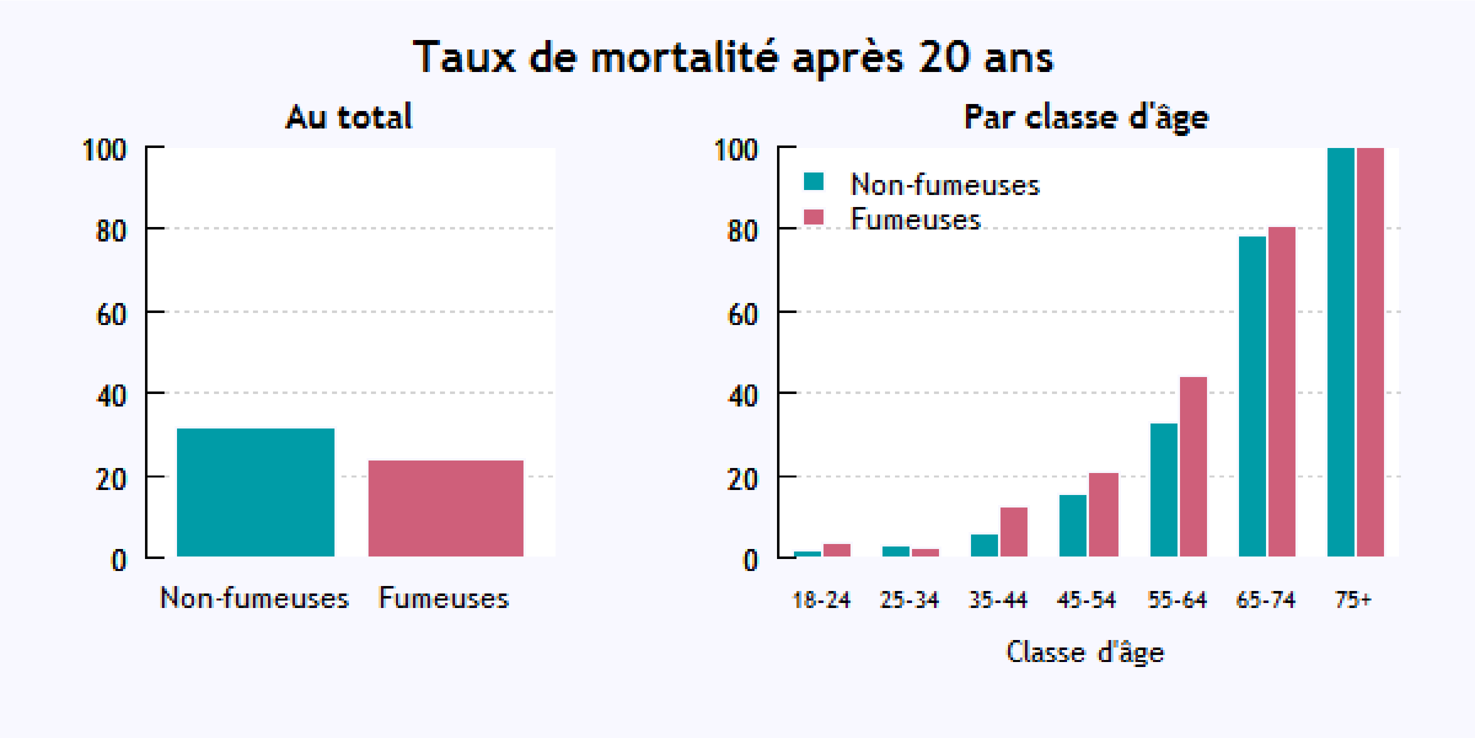
Voilà un modèle du paradoxe de Simpson !

Fumer c’est bon pour la santé !

Une étude ayant pour objectif de comparer le taux de mortalité des fumeuses et des non-fumeuses a été réalisée. Dans cette étude, 1314 femmes ont été suivies pendant 20 ans.

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

Si nous examinons en précision les chiffres, dans l’étude, il y avait 582 fumeuses et 139 sont mortes (cela fait bien 24%), ainsi que 732 non-fumeuses dont 230 sont mortes (31%, pas de problème). Mais si on représentait les chiffres en séparant par classe d’âge ? C’est ce que montre le graphique dans la diapo suivante.



Comme on peut le voir, si on raisonne par classe d’âge, dans chaque tranche la mortalité chez les fumeuses a été supérieure à celle des non-fumeuses.

Mais voyez-vous ce qui cloche ? Il y a plus de femmes âgées chez les non-fumeuses que chez les fumeuses. Et 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 !

Lisa et Bart

Lisa et Bart sont chargé d’améliorer les pages du WEB. Nous allons comparer ces résultats sur une période de deux semaines. Lors la première semaine Lisa améliore 60% des articles qu'elle modifie contre 90 % des articles amélioré par Bart. La deuxième semaine, Lisa n'améliore que 10 % alors que Bart tient un score de 30% . Comme on peut le voir, dans les deux cas Bart l’emporte. Mais une fois les pourcentages combiné, on remarque que Lisa a un plus grand pourcentage que Bart :
Elle a amélioré un peu plus de 55 % des articles qu'elle a modifiés alors que Bart n'en a amélioré même pas 36% .

Comment est-ce possible que ce soit Lisa qui l’emporte alors que les premiers résultats montre que c’est Bart qui l’emporte ?

Le résultat ne peut pas être déduit qu’avec des pourcentages car il dépend également du nombre de page qu’ils ont modifié.
La première semaine Lisa modifie 100 articles, dont 60 articles qu’elle améliore. Alors que Bart s’occupe de 10 articles, dont 9 qu’il modifie. Tandis que la deuxième semaine, elle n’améliore qu’une page sur 10 articles. A l’inverse, lui modifie 100 articles dont 30 améliorés.
Quand nous combinons les deux résultats on remarque que tout les deux ont changé le même nombre de pages (110 pages), mais Lisa en a amélioré 61 contre 39 de la part de Bart. Voici un tableau qui représente la situation.

	SEMAINE 1	SEMAINE 2	TOTAL
LISA	60/100 = 60 %	1/10 = 10 %	61/110 = 55,45 %
BART	9/10 = 90 %	30/100 = 30 %	39/110 = 35,45 %

D’une manière plus simplifiée :

- La première semaine :

  $S_A(1) = 60\ \%$ → Lisa améliore 60 % des articles qu'elle modifie
  $S_B(1) = 90\ \%$ → Bart améliore 90 % des articles qu'il modifie

  -Bart a le meilleur score.

- La deuxième semaine :

  $S_A(2) = 10\ \%$ → Lisa améliore 10 % des articles qu’elle modifie
  $S_B(2) = 30\ \%$ → Bart améliore 30 % des articles qu’il modifie

  -Bart possède encore le meilleur score.

Dans les deux cas, Bart a le meilleur score d’amélioration. Mais si nous combinons les résultats, nous remarquons que Lisa et Bart ont modifié 110 articles. On établit cela ainsi :

$S_A = 61/110$ → Lisa a amélioré 61 articles
$S_B = 39/110$ → Bart en a amélioré seulement 39
$S_A > S_B$ → Lisa repasse en tête

On voit donc que Bart est meilleur pour chaque semaine mais que globalement il est plus mauvais, d’où le paradoxe.

Comment créer le paradoxe de Simpson ?

Après avoir expliqué le paradoxe de Simpson et l’avoir illustré par des exemples, voyons voir comment on peut le créer !

Premièrement, regardons comment s’énonce ce paradoxe.

Ce paradoxe est une corrélation qui peut s’inverser ou disparaître lorsqu’on considère les données dans leurs ensembles.

Les ingrédients

- On a besoin d’un facteur de confusion qui possède une variable qui va changer le résultat final et qui n’est pas vraiment dite au départ. Dans le premier exemple, il s’agissait de la taille des calculs, car celle-ci influe sur la probabilité de succès du traitement. Pour le deuxième exemple il s’agissait de l’âge des personnes, dans lequel se joue la mortalité.

- Ensuite, il faut que les statistiques étudiées ne soit distribuées de la même façon : comme avec le cas du tabac, on trouve plus de vieilles femmes du coté des non-fumeuses que chez les fumeuses. Pour le cas des reins, le traitement A est plus souvent donné sur les gros calculs, et le B sur les petits.

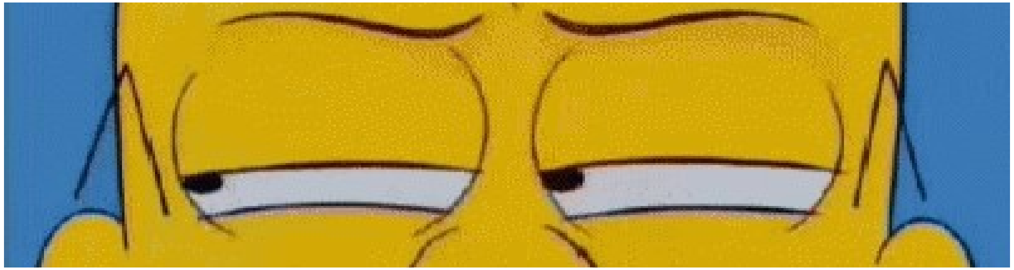
On trouve le potentiel de manipulation qui se cache derrière ce paradoxe : on peut nous faire croire à quelque chose (tel en médecine, ce traitement marche mieux que l’autre...etc.). Mais si nous regardons les chiffres en détail, ces effets peuvent disparaître ou s’inverser !

Alors que faire ?

Comment s’en protéger ?

- Rappelons nous, ce paradoxe existe grâce à une corrélation cachée et influente, et que l’échantillon sur lequel on se base n’est pas le même. C’est pour cela qu’en science on préfère utiliser des expériences « randomisées », qui permettent d’assurer une bonne 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.

- il faut avoir l’œil critique et être méfiant quand ces chiffres viennent de données analysées à posteriori, plutôt que sur des analyses expérimentales qu’on a construit soi-même a priori et en randomisant. Dire 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.



Pour résumer tout cela, ce paradoxe se produit quand il y a une variable cachée et influente. Cela signifie que les chiffres ont peu de sens, et doivent être critiqués par un expert du domaine, capable de détecter l’existence d’un tel facteur.

A l’heure où la mode du « fact-checking » prend de l’ampleur, on a un peu tendance à nous faire croire que les chiffres seraient la vérité absolue. Mais non, la vérité absolue n’existe pas, et on aura toujours besoin de gens au courant pour interpréter correctement des chiffres, qu’ils soient scientifiques, économiques ou médicaux.

SOURCES

- Science étonnante
- Wikipédia