



Esprit critique et information chiffrée : le paradoxe de Simpson

publié le 27/05/2024 - mis à jour le 13/06/2024

TraAM 2023-2024

Descriptif :

Cet article rend compte d'une expérimentation réalisée en classe de seconde sur le paradoxe de Simpson, l'objectif étant de sensibiliser les élèves à certaines situations contre-intuitives sur les proportions.

Sommaire :

- Fiche synoptique
- Présentation du paradoxe de Simpson
- Émergence du phénomène au sein de la classe : travail de groupes
- Reproduction de phénomènes de Simpson par des manipulations vectorielles sous GeoGebra
- Estimation de la fréquence du phénomène par simulation informatique
- Conclusion
- Bonus : une représentation graphique du paradoxe de Simpson avec Python

● Fiche synoptique

○ Thématique

Information chiffrée : le paradoxe de Simpson

○ Niveau concerné

Seconde

○ Problématique

Comment définir et formaliser le paradoxe de Simpson ? Comment le reproduire ? Quelle est sa fréquence ? Est-il explicable ?

○ Contenu

- information chiffrée
- échantillonnage
- vecteurs
- algorithmique et programmation Python (fonctions, boucles, module random)


○ Nombre d'heures utilisées

- travail de groupe pour faire émerger le paradoxe : 1 heure
- formalisation et manipulations vectorielles sur GeoGebra pour reproduire le paradoxe : 1 heure
- simulations en salle informatique pour estimer la fréquence de survenue du paradoxe : 1 heure
- reconstitution d'une situation menant au paradoxe de Simpson et bilan : 1 heure

○ Compétences du CRCN travaillées

- 1. Information et données :
 - 1.2 Gérer des données ;
 - 1.3 Traiter des données
- 3. Création de contenus :
 - 3.4 programmer ;
- 5. Environnement numérique :
 - 5.2 Évoluer dans un environnement numérique ;

Outils et ressources

- calculatrice
- GeoGebra
- Plateforme Capytale pour la programmation Python
- [Article de Jean-Paul Delahaye publié en juillet 2013 dans la revue *Pour la science*](#) ↗
- Ensemble des documents utilisés pour la séquence :
 -  [Ensemble des ressources professeurs pour l'expérimentation du paradoxe de Simpson](#) (Zip de 4.9 Mo)
Ressources (fichiers latex, GeoGebra, Python...) utilisées pour l'expérimentation

► Page suivante : "*Présentation du paradoxe de Simpson*"

● Présentation du paradoxe de Simpson

Le paradoxe de Simpson en statistiques désigne une situation où une tendance observée au sein de plusieurs groupes s'inverse lorsque ces groupes sont regroupés.

Afin d'illustrer ce paradoxe, nous reprenons l'exemple proposé par Jean-Paul Delahaye dans [son article publié en juillet 2013 dans la revue *Pour la science*](#) ↗.

On mène des tests en double aveugle sur un nouveau médicament traitant une maladie.

On a traité 160 patients (80 hommes et 80 femmes) dont 80 ont reçu le médicament, et les 80 autres un placebo. On obtient les statistiques suivantes :

- Taux de guérison dans le groupe des femmes :

Femmes	Guéries	Non guéries	taux de guérison
Médicament	4	16	20 %
Placebo	18	42	30 %

On peut en conclure que, parmi les femmes, le placebo réussit mieux que le médicament.

- Taux de guérison dans le groupe des hommes :

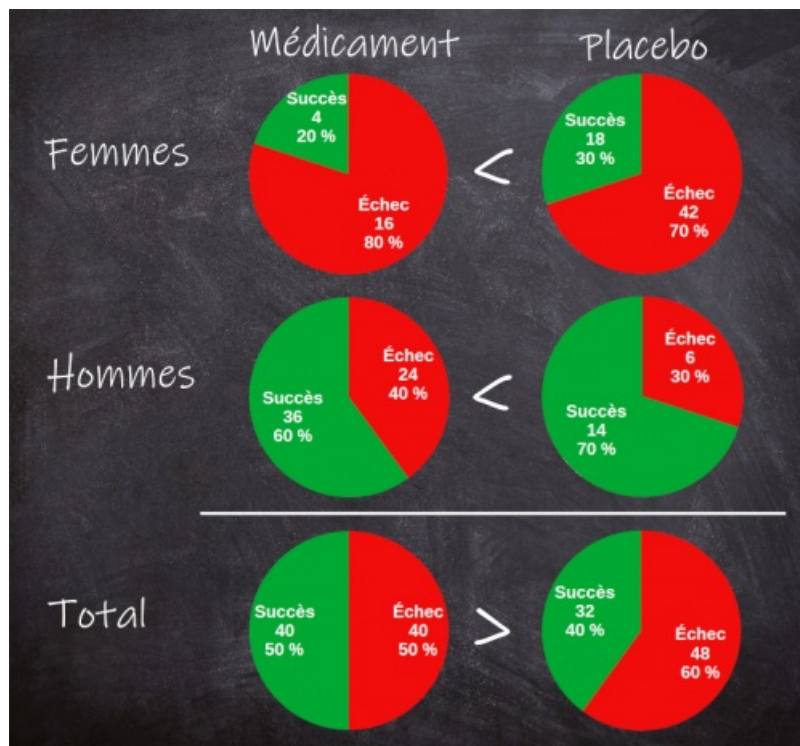
Hommes	Guéris	Non guéris	taux de guérison
Médicament	36	24	60 %
Placebo	14	6	70 %

On peut en conclure que, parmi les hommes, le placebo réussit mieux que le médicament.

- Taux de guérison dans l'ensemble du groupe :

Ensemble	Guéries	Non guéries	taux de guérison
Médicament	40	40	50 %
Placebo	32	48	40 %

En regroupant hommes et femmes, le médicament produit de meilleurs résultats que le placebo, ce qui va à l'encontre des conclusions dans les deux groupes séparés.



Le médicament est moins efficace que le placebo chez les hommes ainsi que chez les femmes. Or en regroupant les hommes et les femmes, le médicament semble plus efficace que le placebo.

On se trouve dans un cas du paradoxe de Simpson : la fusion de données concluant individuellement dans un sens – l'inutilité du médicament testé – donne des résultats concluant dans le sens inverse, l'utilité du médicament.

La première occurrence de ce phénomène dans la littérature scientifique remonte à 1899, lorsque le mathématicien anglais Karl Pearson a décrit des données similaires. Plus tard, en 1903, le statisticien écossais Undy Yule a redécouvert ce phénomène, suivi par le statisticien britannique Edward Simpson qui a publié en 1951 un article détaillé étudiant et discutant cette singularité statistique, laissant son nom à ce paradoxe.



Les statistiques vues de biais : le paradoxe de Simpson | Voyages au pays des maths | ARTE (Video Youtube)
 Les statistiques vues de biais : le paradoxe de Simpson | Voyages au pays des maths | ARTE

L'expérimentation imaginée à partir de ce phénomène avait pour but de questionner les élèves sur leurs représentations des proportions et sur leur sens critique vis-à-vis de phénomènes chiffrés surprenants, notamment dans le cas d'une prise de décision : dans l'exemple étudié, faut-il, oui ou non prescrire ce médicament ?

Les séances en classe se sont déroulées en plusieurs temps :

- émergence du paradoxe de Simpson auprès des élèves par une étude en groupe de situations paradoxales ;
- reproduction de phénomènes de Simpson à partir de considérations vectorielles et rédaction de situations s'y

rapportant ;

- mesure de la fréquence du paradoxe par simulation informatique
- recherche documentaire pour expliquer le phénomène

► Page suivante : "Émergence du phénomène au sein de la classe : travail de groupes"

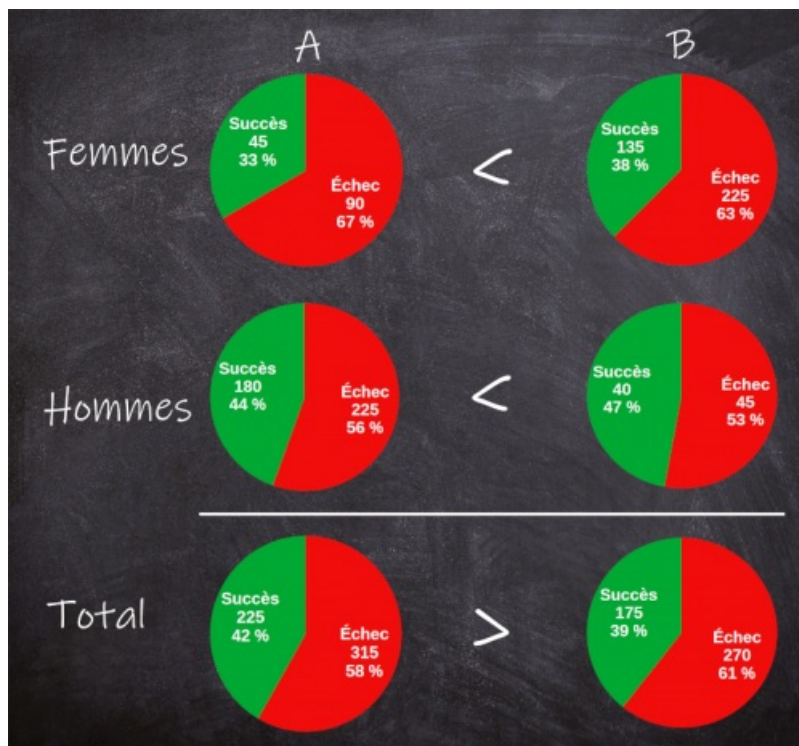
● Émergence du phénomène au sein de la classe : travail de groupes

La première séance s'est déroulée au retour des vacances de printemps. Après une brève présentation de l'activité, les élèves se sont répartis en groupes de 3 ou 4 et se sont vu attribuer une situation où ils devaient calculer la fréquence de survenue d'un événement (guérison, réussite, vote favorable,...) dans deux catégories d'une première sous-population, puis d'une deuxième et enfin dans le regroupement des deux-sous populations.

 [Énoncé des situations pour le travail de groupe](#) (PDF de 105.7 ko)

Énoncé des situations pour le travail de groupe servant d'introduction au paradoxe de Simpson

Les huit situations proposées étaient tirées de cas réels et menaient toute à un paradoxe de Simpson : l'ordre des fréquences des deux catégories était le même dans chacune des deux-sous populations mais il s'inversait lorsqu'on regroupait les deux-sous populations.



Comparaison de la réussite de deux anti-migraineux A et B, chez les hommes et chez les femmes, puis dans la population totale

Après une vingtaine de minutes de recherche en groupes, les calculs de fréquences ont été faits, avec parfois des erreurs sur la population de référence traduisant une compréhension partielle de la situation. Les résultats ont rarement suscité des réactions dans les groupes et il a fallu que l'enseignant explicite les valeurs obtenues ainsi que leur ordre pour que les élèves prennent conscience du paradoxe. En revanche, une fois mis en évidence, le phénomène a réellement interpellé leur intuition et ils ont eu le réflexe de chercher une cause dont la première qui leur est venue à l'idée : ils s'étaient trompés dans leurs calculs.

Un bilan en classe entière où chaque groupe a rendu compte de sa situation et de ses résultats a permis de rassurer les groupes sur leurs capacités calculatoires : il n'y avait pas d'erreur de calcul et tous les groupes ont obtenu la même conclusion surprenante qui restait valable sur les fréquences comme sur les ratios.

Le paradoxe a été nommé et l'enseignant n'a pas fait de commentaire ni entamé de tentative d'explication du phénomène afin de maintenir le mystère entourant cette singularité.

Pour conclure cette première séance et dans le but de préparer la suite de l'expérimentation, une formalisation a été proposée en classe entière : elle a permis de montrer aux élèves le point commun des diverses situations et qu'elles relevaient toutes d'un même phénomène qui pouvait être structuré sous la forme de comparaison de fractions.

► Page suivante : "Reproduction de phénomènes de Simpson à l'aide GeoGebra"

● Reproduction de phénomènes de Simpson par des manipulations vectorielles sous GeoGebra

Lors de la deuxième séance, la classe a ensuite été orientée vers un objectif de reproduction du phénomène dans le but d'appréhender sa fréquence de survenue :

Le paradoxe de Simpson est-il fréquent ou plutôt rare ? Est-il facile ou non de le reproduire ? Comment l'obtenir ?

Il paraissait risqué de les laisser tâtonner à la recherche de 8 valeurs menant à un paradoxe de Simpson.

L'idée suggérée a alors été de limiter la recherche pour des nombres entiers entre 1 et 5 avec une traduction géométrique de la situation, les ratios étant considérés comme des "pentes" de vecteurs. Ce passage du cadre numérique au cadre graphique a permis de réinvestir la notion de pente abordée dans le précédent chapitre. En effet, les 4 vecteurs construits avaient pour coordonnées les 8 paramètres et leur "pente" correspondaient aux quatre ratios de départ.

Il était donc plus facile et plus rapide de comparer les ratios en comparant les pentes des vecteurs associées, par simple contrôle visuel.

Il restait ensuite à représenter la réunion des deux populations et la formalisation précédente a naturellement guidé les élèves vers la somme des vecteurs.

Ainsi, en deux instructions GeoGebra, ils ont obtenu deux vecteurs représentant les ratios dans la population totale. Il restait ensuite, par des manipulations à la souris, à déplacer les extrémités des vecteurs de départ pour obtenir un paradoxe de Simpson :

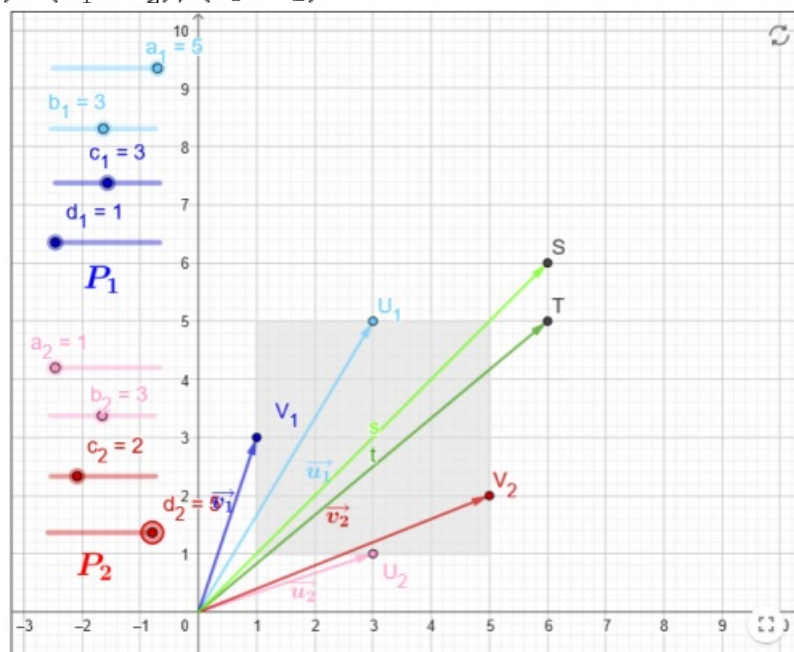


Fichier GeoGebra de manipulation de vecteurs (Geogebra Tube)

Fichier GeoGebra de manipulation des vecteurs : il y aura paradoxe de Simpson lorsque la "pente" de v_1 sera supérieure à celle de u_1 et que celle de v_2 sera supérieure à celle de u_2 mais que celle de t sera inférieure à celle de s

Les élèves ont donc tâtonné pendant un bon quart d'heure pour obtenir des positions satisfaisantes :

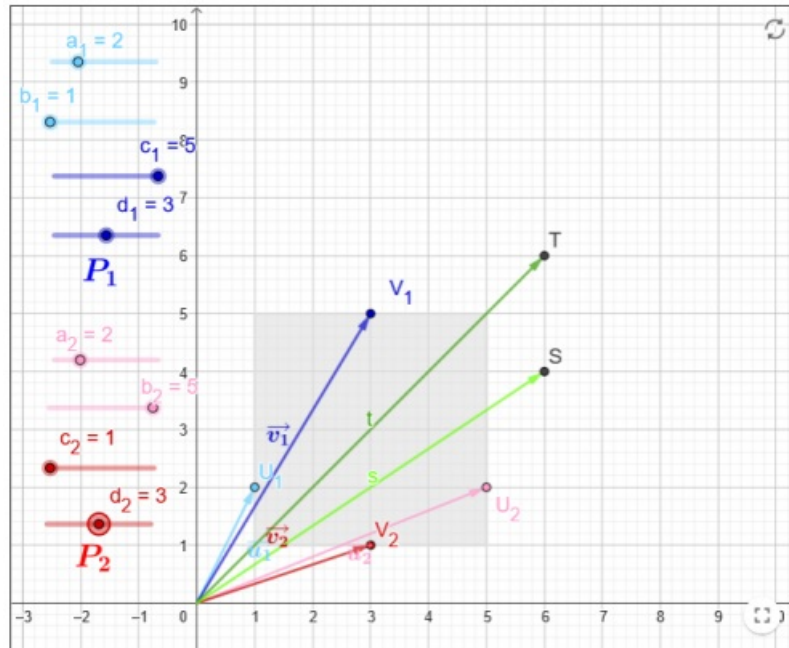
- Exemple de paradoxe de Simpson de type 1 : le vecteur bleu foncé a une "pente" supérieure à celle du bleu clair $c_1/d_1 > a_1/b_1$, le vecteur rouge foncé a une "pente" supérieure à celle du rouge clair $c_2/d_2 > a_2/b_2$ mais le vecteur vert foncé a une "pente" inférieure à celle du vert clair $(c_1+c_2)/(d_1+d_2) < (a_1+a_2)/(b_1+b_2)$:



Exemple de paradoxe de Simpson de type 1 obtenu à l'aide de GeoGebra

- Exemple de paradoxe de Simpson de type 2 : le vecteur bleu foncé a une "pente" inférieure à celle du bleu clair $c_1/d_1 < a_1/b_1$, le vecteur rouge foncé a une "pente" inférieure à celle du rouge clair

$c_2/d_2 < a_2/b_2$ mais le vecteur vert foncé a une "pente" supérieure à celle du vert clair
 $(c_1+c_2)/(d_1+d_2) > (a_1+a_2)/(b_1+b_2)$:



Exemple de paradoxe de Simpson de type 2 obtenu à l'aide de GeoGebra

En guise de conclusion, la séance s'est terminée sur le constat empirique que le paradoxe de Simpson est reproductible mais, en terme de fréquence, il s'agit d'un phénomène qui semble plutôt rare étant donnée la difficulté d'obtenir une configuration graphique caractéristique du paradoxe de Simpson.

Logiquement, il a ensuite été proposé aux élèves d'estimer cette fréquence à l'aide d'une simulation informatique.

● Estimation de la fréquence du phénomène par simulation informatique

Lors de la troisième et quatrième séance, la classe s'est partagée en deux et a effectué tour à tour les deux activités suivantes :

- simuler le tirage de 8 entiers aléatoires dans un intervalle et comptabiliser la survenue du paradoxe de Simpson afin d'estimer sa fréquence "naturelle" ;
- réutiliser les données obtenues avec GeoGebra pour imaginer un énoncé menant à un paradoxe de Simpson afin de travailler la recontextualisation du phénomène

► Page suivante : "*Simulation informatique*"

○ Simulation informatique

Le code proposé se décompose en une fonction de test de paradoxe de Simpson et une fonction de simulation. Il a été complété en classe entière car les élèves disposent de peu de recul sur la programmation Python. Néanmoins, ce travail collectif a été l'occasion de réactiver les structures fondamentales de l'algorithmique déjà travaillées lors de [l'expérimentation sur le problème des partis](#) :

- notion de fonction informatique ;
- utilisation de structures conditionnelles et de variables booléennes pour caractériser le paradoxe de Simpson
- utilisation des fonctions aléatoires de Python pour simuler le choix aléatoire d'un entier dans un intervalle.
- utilisation d'un compteur et d'une boucle pour réaliser un échantillon de tirages aléatoires et estimer une fréquence de survenue du paradoxe de Simpson

```

1. import random # on importe le module random pour générer des nombres aléatoires
2.
3. # Remplacer les None par votre code et exécuter le script
4. def test_paradoxe_simpson(a1,b1,c1,d1,a2,b2,c2,d2):
5.     """renvoie True si les 8 paramètres donnent un paradoxe de Simpson et False sinon"""
6.     if None: # paradoxe de Simpson de type 1
7.         return None
8.     elif None: # paradoxe de Simpson de type 2
9.         return None
10.    else: # dans les autres cas
11.        return None
12.
13. # test de la fonction test_paradoxe_simpson : vous devez avoir True dans la console
14. print(test_paradoxe_simpson(3,1,4,1,1,2,3,5) == True)
15.
16. # tester avec les valeurs que vous avez relevées à l'aide de GeoGebra
17.
18. # Remplacer les None par votre code et exécuter le script
19. def frequence_simpson(taille_intervalle, nb_repetitions):
20.     """renvoie la fréquence d'obtention du paradoxe de Simpson pour un échantillon de taille nb_repetitions de 8 nombres aléatoires pris dans l'intervalle [1,taille_intervalle]"""
21.     compteur = 0
22.     for tirage in range(None):
23.         a1= random.randint(1, taille_intervalle)
24.         b1= random.randint(1, taille_intervalle)
25.         c1= random.randint(1, taille_intervalle)
26.         d1= random.randint(1, taille_intervalle)
27.         a2= random.randint(1, taille_intervalle)
28.         b2= random.randint(1, taille_intervalle)
29.         c2= random.randint(1, taille_intervalle)
30.         d2= random.randint(1, taille_intervalle)
31.         if None == True:
32.             compteur= None

```

Les élèves devaient ensuite compléter un script à trous sur Capytale et appeler la fonction avec différentes valeurs de la taille de l'intervalle de tirage et différentes tailles d'échantillon.

Lors des appels pour des échantillons à tirage entre 1 et 5, notamment `frequence_simpson(5,100)`, il n'a pas été rare d'obtenir une fréquence 0 et il a fallu passer à des tailles d'échantillons plus importantes pour obtenir des fréquences non nulles qui s'approchaient de 0,00119 pour des échantillons de taille un million. Cette valeur a été mise en avant lors du bilan car elle correspondait à la situation de GeoGebra dans laquelle la probabilité théorique de tomber au hasard sur un paradoxe de Simpson était de $\frac{464}{5^8} \approx 0,001188$. Cela a permis aux élèves de quantifier la rareté du phénomène et de mieux comprendre pourquoi ils avaient eu autant de mal à trouver un paradoxe de Simpson avec les manipulations GeoGebra.

Le reste des simulations a permis de généraliser la rareté du paradoxe de Simpson et de retrouver la conclusion de l'article de Jean-Paul Delahaye :

La proportion de jeux de huit entiers pris entre 1 et n vérifiant les inégalités du paradoxe (ou les inégalités inverses qui sont aussi paradoxales) tend vers 0,9606 % quand n tend vers l'infini.

taille \ repet.	100	1000	10000	100000	1000000
5	0,0	0,001	0,0015	0,00121	0,001145
100	0,0	0,007	0,0089	0,00925	0,009115
1000	0,0	0,014	0,0086	0,01	0,009431
10000	0,0	0,007	0,0092	0,00919	0,00958
100000	0,1	0,011	0,0109	0,00983	0,009403
1000000	0,04	0,004	0,0096	0,0098	0,009604

Relevé élève de fréquences d'apparition du paradoxe de Simpson par simulation informatique (cliquer sur l'image pour l'agrandir)

 [Relevés de fréquences d'apparition du paradoxe de Simpson par simulation informatique](#) (PDF de 411.9 ko)

Relevés élèves de fréquences d'apparition du paradoxe de Simpson par simulation informatique

- Page suivante : "Création de situations et conclusion"

[O Création de situations menant au paradoxe de Simpson](#)

Cette demande, plutôt inhabituelle, visait à amener les élèves à développer leur capacité à mobiliser, à transférer les

savoirs construits lors des séances précédentes et exercer leur esprit critique pour obtenir une situation crédible. En inventant un énoncé de situation proche de la situation de référence travaillée en classe, l'enseignant peut avoir un retour sur la compréhension de la structure interne de la situation par l'élève. En s'engageant dans la création d'un énoncé, l'élève explicite une démarche et argumente son raisonnement, ce qui nécessite une compréhension et une analyse fine de la structure interne de la situation.

Les données numériques de base étaient les octuplets obtenus lors des manipulations de GeoGebra mais la demande supplémentaire était qu'ils mettent en jeu des effectifs plausibles : les nombres obtenus étant inférieurs à 5, ils ne pouvaient être pris tels quels pour représenter des effectifs de population.

La réflexion s'est alors portée sur les mécanismes calculatoires mis en jeu dans le paradoxe de Simpson : que l'on parle de fréquence ou de ratios, l'opération arithmétique centrale était la division. Or le quotient de deux nombres ne change pas lorsqu'on multiplie ces deux nombres par un même nombre non nul. Au travers de ce rappel et des échanges avec les élèves, l'enseignant les a progressivement amenés à comprendre que le paradoxe de Simpson était conservé si on multipliait tous les éléments de l'octuplet par un même facteur non nul.

Dégagés de cette contrainte numérique, il leur a fallu réfléchir aux "ingrédients" de la recette : deux sous-populations, deux catégories dans ces populations, un critère binaire...

Pour achever leur production écrite, il leur a tout de même été demandé de calculer les ratios ou les fréquences afin de retrouver l'ordre paradoxal des rapports.

Votre situation :
Des chercheurs de Brébré ont publié une étude sur l'efficacité pour éliminer les crampes.
Traitement A : pommade
Traitement B : comprimé
la population étudiée a d'abord été divisée selon deux groupes :
→ groupe Femme
→ groupe Homme
On obtient les résultats pour ces deux groupes :

P_1 H	R	\bar{R}
Traitement A	300	100
Traitement B	500	200

P_2 F	R	\bar{R}
T: A	300	500
T: B	200	400

Lorsqu'on regroupe les deux populations :

$P_1 + P_2$	\bar{R}	\bar{R}
T: A	600	600
T: B	700	600

2. Calculer les six ratios qui permettent de mettre en évidence un paradoxe de Simpson :

- Ratio R/\bar{R} de C_1 dans P_1 : $\frac{300}{400} = 0,75$
- Ratio R/\bar{R} de C_2 dans P_1 : $\frac{500}{700} = 0,71$
- Ratio R/\bar{R} de C_1 dans P_2 : $\frac{300}{500} = 0,6$
- Ratio R/\bar{R} de C_2 dans P_2 : $\frac{200}{400} = 0,5$

Puis dans la population totale $P_1 + P_2$:

- Ratio R/\bar{R} de C_1 dans $P_1 + P_2$: $\frac{300+300}{400+500} = 0,6$
- Ratio R/\bar{R} de C_2 dans $P_1 + P_2$: $\frac{500+200}{700+600} = 0,5$

3. Refaire les calculs avec les fréquences.

$H \rightarrow R$ $TA : \frac{300}{400} = 0,75$ $TB : \frac{500}{700} = 0,71$	$P \rightarrow R$ $TA : \frac{300}{800} = 0,375$ $TB : \frac{200}{500} = 0,4$	$H + F \rightarrow R$ $TA : \frac{600}{1100} = 0,545$ $TB : \frac{200}{1300} = 0,154$
---	---	---

Un exemple de production d'énoncé sur le paradoxe de Simpson par un.e élève (cliquer sur l'image pour l'agrandir)

[Inventions d'énoncés d'élèves pour illustrer le paradoxe de Simpson](#) (PDF de 1012.2 ko)
Inventions d'énoncés d'élèves pour illustrer le paradoxe de Simpson

Enfin, pour terminer l'expérimentation, il leur a été demandé de faire des recherches complémentaires sur Internet afin de déterminer des facteurs explicatifs de survenue du paradoxe de Simpson lors d'études statistiques réelles dans des domaines aussi variés que la médecine, l'éducation, le sport...

Au-delà de l'existence "naturelle" du paradoxe dans des données aléatoires, les élèves ont pu obtenir, grâce à ces recherches documentaires, des éclaircissements liés au concept de **variable de confusion**.

Une variable de confusion (ou facteur de confusion) est une variable qui est liée à la fois à la variable dépendante

(ou variable de résultat) et à la variable indépendante (ou variable d'exposition) dans une étude statistique ou épidémiologique. Elle peut masquer, atténuer ou même inverser la relation entre les deux variables d'intérêt.

L'exemple de la [situation n°4](#) sur le lien entre mortalité et tabagisme sur un groupe de femmes dans les années 70 a alors été explicité en classe pour mettre en avant l'âge des femmes comme facteur de confusion.

Les femmes âgées, qui ont un taux de mortalité plus élevé, sont sur-représentées parmi les non-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.

La variable de confusion (l'âge dans cet exemple) influence à la fois le statut de fumeur/non-fumeur (variable d'exposition) et la mortalité (variable de résultat). Ne pas tenir compte de cette variable peut conduire à des conclusions incorrectes sur la relation entre le tabagisme et la longévité.

● Conclusion

Cette expérimentation avait pour objectif d'interroger l'esprit critique des élèves sur des situations d'information chiffrée contre-intuitives : l'ordre des fréquences de deux catégories dans deux sous-populations n'est pas nécessairement conservé lorsque les deux sous-populations sont réunies. Le paradoxe de Simpson permet de mettre en exergue un biais cognitif lié à l'agrégation des données qui peut produire des résultats contradictoires si on ne tient pas compte des différences de taille des groupes ou d'éventuelles interactions cachées entre certaines variables du problème (facteur de confusion).

Cette expérimentation a plu aux élèves car les données de départ sont concrètes et les conclusions sont accessibles par des calculs simples. Il est intéressant de noter que, face à ces résultats contradictoires, plusieurs groupes ont émis l'hypothèse d'erreurs dans leurs calculs, ce qui a permis de mettre en évidence un nouveau biais cognitif : le biais de confirmation qui incite une personne à douter de ses propres conclusions ou de ses capacités analytiques si ces résultats remettent en question ses convictions préexistantes ou ses attentes. Plutôt que de considérer objectivement les données contradictoires comme des opportunités d'apprentissage ou de réévaluation, elle pourrait interpréter ces résultats comme une indication de son incompetence ou de ses erreurs.

Dans les interactions avec les élèves au moment du travail de groupe, un travail spécifique a été mené pour surmonter ce biais. Par le jeu de questions/réponses sur leur métacognition, l'enseignant a cherché à convaincre les élèves de la validité de leurs procédures afin de renforcer leur confiance en soi mais aussi développer leur sens critique face à des résultats paradoxaux.



[Ensemble des ressources professeurs pour l'expérimentation du paradoxe de Simpson](#) (Zip de 4.9 Mo)

Ressources (fichiers latex, GeoGebra, Python...) utilisées pour l'expérimentation

► Page suivante : "*Bonus*"

● Bonus : une représentation graphique du paradoxe de Simpson avec Python

Cette [représentation graphique du paradoxe de Simpson](#) [↗](#) a été imaginée par Jean-François Colonna, chercheur au chercheur au centre de mathématiques appliquées (CMAP) de l'École Polytechnique.

Dans le cas de notre étude avec GeoGebra, on a envisagé tous les cas lorsque les huit éléments des deux tableaux parcourent toutes les valeurs entières entre 1 et 5 ce qui fait $5^8 = 390625$ cas différents.

Chacun de ces 390625 cas est représenté par un pixel dans une image carrée de côté $5^4 = 625$ pixels.

Chaque pixel est repéré par son adresse correspondant à son indice de ligne et son indice de colonne. Ces deux coordonnées vont fournir les 8 nombres nécessaires :

- les quatre coefficients du premier tableau sont obtenus en prenant les quatre chiffres de l'écriture de l'indice de ligne en base 4, que l'on augmente chacun d'une unité pour obtenir des nombres entre 1 et 5 ;
- de même les quatre coefficients du second tableau sont obtenus en prenant les quatre chiffres de l'écriture de l'indice de colonne en base 4, que l'on augmente chacun d'une unité pour obtenir des nombres entre 1 et 5 ;

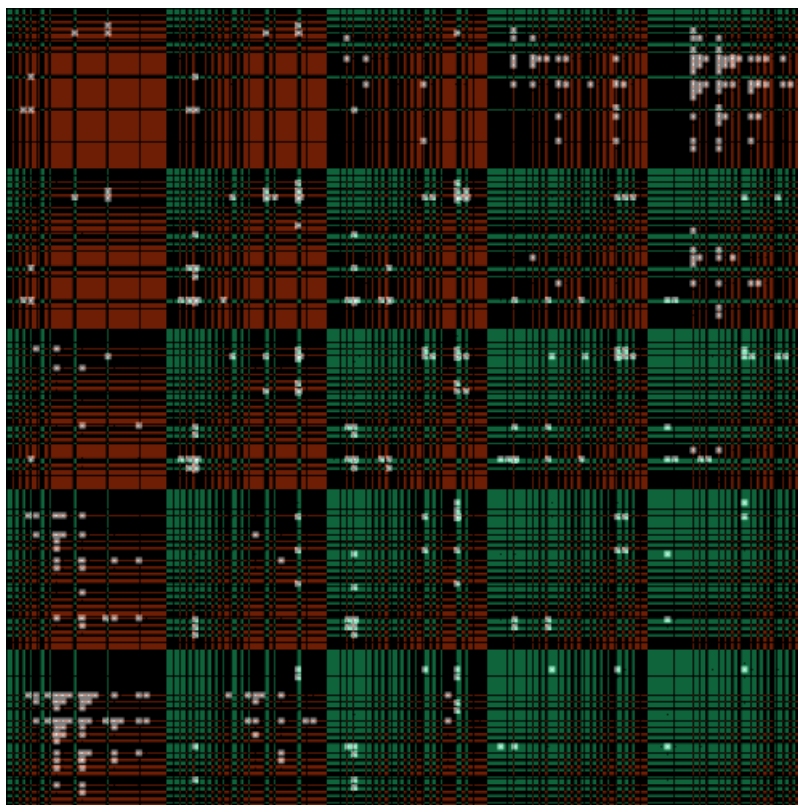
Ensuite chaque pixel est colorié de la manière suivante :

- Si on a un cas « normal » (les tableaux 1 et 2 donnent une même conclusion, qui est aussi celle du tableau de regroupement), le pixel est vert ou marron selon l'ordre des catégories ;
- Si on a un cas sans grand intérêt (les tableaux 1 et 2 se contredisent, ou il y a égalité entre les succès de chaque catégorie), le pixel est noir.
- Si on a un cas de paradoxe de Simpson (le tableau de regroupement s'oppose aux tableaux 1 et 2 qui vont dans le même sens), le pixel est blanc, entouré d'un halo pour aider à le repérer.

Le programme Python téléchargeable ci-dessous implémente cette représentation pour une image paramétrable (taille de 2 à 15) :

 [Programme Python pour une représentation graphique du paradoxe de Simpson](#) (Zip de 2 ko)
Programme Python pour une représentation graphique du paradoxe de Simpson

Et on obtient l'image suivante dans le cas d'un appel avec la taille 5 :



Représentation graphique du paradoxe de Simpson pour des valeurs aléatoires entre 1 et 5

Cette représentation illustre de manière éclatante la rareté du paradoxe de Simpson dans le cas de tirages aléatoires : 464 pixels blancs sur 390 625.

 [Ensemble des ressources professeurs pour l'expérimentation du paradoxe de Simpson](#) (Zip de 4.9 Mo)
Ressources (fichiers latex, GeoGebra, Python...) utilisées pour l'expérimentation