

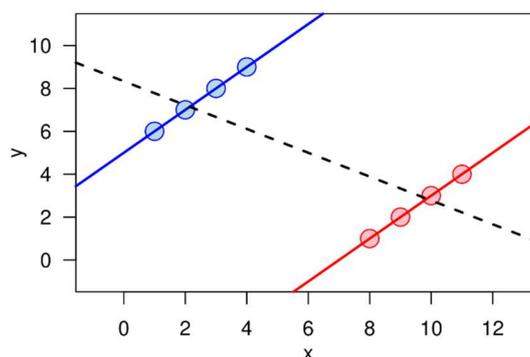
2. Le paradoxe de (Yule) Simpson

Approfondissement de la notion de facteur de confusion. Exemples avec détails mathématiques. Découverte du paradoxe de Simpson et exemples.

#B2D - Travaux dirigés. Prérequis : aucun.

Apprendre

Le « paradoxe de Simpson » est un paradoxe statistique. Il se produit lors de l'observation et l'analyse d'une mesure sur une population donnée. Comme illustré sur le graphique ci-contre, chaque groupe de la population, le groupe bleu et le groupe rouge, semble montrer une corrélation positive entre la mesure des grandeurs x et y . Cependant, si l'on observe les deux groupes comme un seul ensemble, la tendance générale montre une corrélation négative entre x et y . Il n'est pas toujours facile d'identifier les groupes composants la population totale.



Mais alors, que faut-il conclure ? Les variables d'intérêt x et y sont-elles en corrélation positive ou négative ? C'est le paradoxe décrit par George Udny Yule (1903) et Edward Simpson (1951).

Appliquer

Voici un tableau de données statistiques concernant les réussites au BAC d'un ensemble d'élèves issus de 2 groupes :

	2022		2023	
	Inscrits	Admis	Inscrits	Admis
Groupe 1	22	12	15	8
Groupe 2	3	3	10	9

1. Compléter le tableau avec les totaux d'inscrits et d'admis par année.
2. Calculer le pourcentage d'admission global pour chaque année. Conclure.
3. Calculer le pourcentage d'admission par groupe et par année. Conclure.
4. Que peut-on conclure sur l'évolution de la performance entre les deux années ?

Analyser

Dans l'exemple précédent, afin de pouvoir tirer une conclusion sur les performances générales des élèves, il faut regarder en détail l'évolution des groupes.

Le groupe 1 est passé de 55% à 53% de réussite. Le groupe 2 lui, est passé de 100% à 90% de réussite.

On constate en premier lieu que le groupe 1 réussit de manière générale beaucoup moins bien que le groupe 2. On remarque également que le groupe 1 est sur-représenté, surtout pour l'année 2022.

Pour l'année 2023, le nombre d'élèves dans le groupe 2 a considérablement augmenté. En conséquence, même si le groupe 2 a perdu en performance, sa moyenne générale bien plus élevée et son augmentation d'effectif fait pencher la moyenne générale sur une augmentation.

« Les chiffres parlent d'eux-mêmes ! » Cette expression semble alors complètement fautive. Les chiffres ne disent absolument rien. Il est obligatoire de comprendre le **contexte** et de les **interpréter**. Conclure trop rapidement mène bien souvent à des résultats complètement erronés.

Appliquer ★

Reprendre les mêmes consignes que l'exercice précédent, avec cette fois-ci un tableau de données concernant un vaccin et des groupes de personnes avec des facteurs de risque différents :

	Vaccinés		Non vaccinés	
	Formes bénignes	Formes graves	Formes bénignes	Formes graves
Groupe 1 (personnes robustes)	5	0	44	1
Groupe 2 (personnes faibles)	39	6	3	2

Que peut-on conclure sur l'efficacité du vaccin ?

Vous pouvez reprendre les questions de l'exemple précédent.