Statistiques à deux variables
Problème de recherche : le palmarès des lycées

Voici les résultats au baccalauréat de deux lycées d'une même ville, selon la voie (générale ou technologique) :

	Lycée A		Lycée B	
	Candidats	Reçus	Candidats	Reçus
Voie générale	300	270	100	92
Voie technologique	100	60	300	195

Un journal local titre : « Le lycée A écrase le lycée B ! »

La mission

1. Calculer le taux de réussite global de chaque lycée. Le titre du journal est-il justifié ?
2. Calculer le taux de réussite de chaque lycée dans chaque voie. Que constate-t-on ?
3. Expliquer comment ces deux constats peuvent être vrais en même temps.

Pour aller plus loin

Proposer une façon plus juste de comparer les deux lycées, par exemple en imaginant qu'ils accueillent la même proportion de candidats dans chaque voie. Quel lycée obtient alors le meilleur résultat ?

Coups de pouce

Bloqué ? Des indices progressifs sont disponibles en ligne. N'en ouvrir qu'un à la fois, et seulement après avoir cherché !

Éléments de correction

1. Taux globaux

Lycée A : $\frac{270+60}{400} = \frac{330}{400} = 82{,}5\,\%$. Lycée B : $\frac{92+195}{400} = \frac{287}{400} = 71{,}75\,\%$.
Globalement, le lycée A a un taux de réussite supérieur de près de 11 points : le titre semble justifié…

2. Taux par voie

	Lycée A	Lycée B
Voie générale	$\frac{270}{300} = 90\,\%$	$\frac{92}{100} = 92\,\%$
Voie technologique	$\frac{60}{100} = 60\,\%$	$\frac{195}{300} = 65\,\%$

Dans chacune des deux voies, c'est le lycée B qui réussit le mieux!

3. L'explication

Dans les deux lycées, la voie générale a un bien meilleur taux de réussite que la voie technologique. Or 75 % des candidats du lycée A sont en voie générale, contre 25 % seulement au lycée B.
Le taux global est une moyenne des taux par voie, pondérée par ces effectifs : celui du lycée A est « tiré vers le haut » par sa forte proportion de candidats en voie générale. Ce phénomène s'appelle le **paradoxe de Simpson** (enrichissement).

Pour aller plus loin

Avec une même répartition, par exemple 50 % de candidats dans chaque voie :
lycée A : $0{,}5 \times 90\,\% + 0{,}5 \times 60\,\% = 75\,\%$; lycée B : $0{,}5 \times 92\,\% + 0{,}5 \times 65\,\% = 78{,}5\,\%$.
À répartition égale, le lycée B obtient le meilleur résultat. C'est ce que fait le ministère avec ses « indicateurs de valeur ajoutée » : il compare le taux de réussite d'un lycée à celui qu'on attendrait compte tenu de ses élèves.

Ce qu'on a utilisé du chapitre

— Lire un tableau croisé d'effectifs.
— Calculer des fréquences marginales et conditionnelles.
— Interpréter et critiquer des indicateurs statistiques.