

D'après les statistiques communiquées par une compagnie d'assurances automobiles :

3) *Les voitures roulant à une vitesse modérée provoquent beaucoup plus d'accidents que celles qui roulent à plus de 180 km/h.*

Doit-on en conclure qu'il est plus dangereux de rouler à vitesse modérée ?

4) *Beaucoup d'accidents arrivent à proximité du domicile des conducteurs.*

Doit-on en conclure qu'un automobiliste est plus en sécurité loin de chez lui que près de chez lui ?

Questions :

Après avoir répondu aux questions, dans quelle(s) classe(s) peut-on poser ce pb ? Le situer dans une progression. A-t-il un intérêt pédagogique ? (Préciser éventuellement lequel).

On parle bien de statistiques descriptives. On peut accepter l'idée que cette étude porte sur un échantillon de taille importante, du moins et sans doute sur la population des assurés de la compagnie d'assurance, les individus (statistiques) étant a priori les conducteurs.

- 1) Les individus (statistiques) pourraient être plutôt les accidents provoqués. La variable : la vitesse (avec deux classes/modalités précisées : « plus de 180 » et « modérée »). Ceci est une modélisation possible.

Le nombre de voitures roulant à vitesse modérée est a priori bien plus important que celui de celles roulant à plus de 180 km/h. Il n'y a donc rien d'étonnant à ce que le nombre d'accidents dans la première catégorie soit plus important. On parle ici d'effectif (terme statistique).

Cette remarque met en évidence l'importance d'une information relative permettant la comparaison : la fréquence de conducteurs de l'étude impliqués dans un accident de chaque cas (plus de 180km/h ou « modérée »). Cela amène à une nouvelle modélisation : les individus (statistiques) étant les trajets, les variables : la vitesse du trajet (« plus de 180 » et « modérée ») et accident (oui/non).

On aurait donc une information du type suivant (avec des données fictives bien sur) qui serait compatible avec l'énoncé.

Vitesse / Accident	Oui	Non	Total
Modérée	10	1980	1990
+ de 180	1	9	10

- 2) C'est la même idée. Pour pouvoir conclure, il faudrait avoir une information sur le nombre de trajets (voire la part de chaque trajet) à proximité du domicile et loin du domicile.

Ces exercices peuvent être posés au le collège dès le début du cycle 4 (mais également au lycée avec peut-être d'autres objectifs). Plusieurs utilisations sont possibles, il n'y a pas de « bonne » réponse ici, on peut par exemple l'envisager en début de séquence pour réfléchir à la notion d'information relative et donc à la notion de fréquence.

Il y a donc un intérêt pédagogique principal : mettre en avant la fréquence comme outil de comparaison. Il permet également une discussion sur la pertinence et la maîtrise d'outils précis en sciences des données permettant une objectivation des inférences possibles.