

Quelques éléments méthodologiques permettant d'analyser des données relationnelles

La majorité des outils que j'ai utilisé font appel au traitement de données. Le piège de cette démarche est d'utiliser des résultats fournis par l'ordinateur sans vérifier la légitimité de cette utilisation. Pour comprendre comment il faut s'y prendre pour analyser des données relationnelles, observons deux situations. Les données sont issues d'une enquête, faite sur les ressentis des élèves de cinq collèges de l'Oise, que nous avons souvent citée dans le roman.
Voir la totalité des résultats à l'adresse :
https://evallib.org/Documentation/5-col-Oise/5-col-Oise.pdf

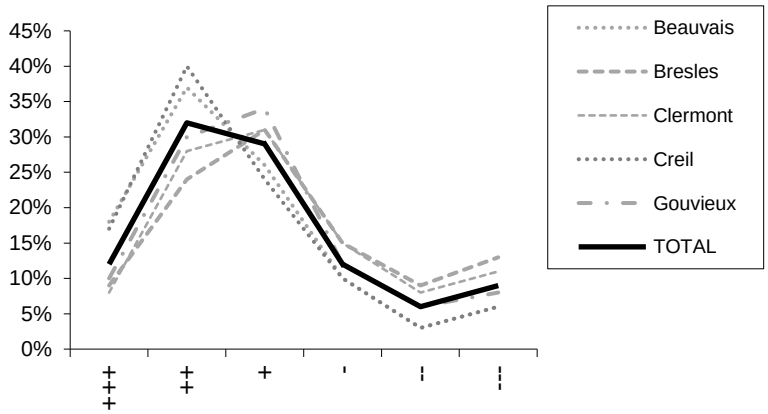
⇒ Première situation

Pour comprendre observons les données suivantes :

Comment juges-tu les conditions de vie au collège ?

	+++	++	+	-	--	---	TOTAL
Beauvais	18%	37%	26%	10%	3%	6%	100%
Bresles	9%	24%	31%	15%	9%	13%	100%
Clermont	8%	28%	31%	15%	8%	11%	100%
Creil	17%	40%	24%	10%	3%	6%	100%
Gouvieux	10%	30%	34%	12%	6%	8%	100%
TOTAL	**12%**	**32%**	**29%**	**12%**	**6%**	**9%**	**100%**

Chaque ligne correspondant à un collège est calculée en pourcentage par rapport à ses propres effectifs.
La ligne « TOTAL » est obtenue en prenant en compte l'ensemble des répondants des cinq établissements puis en calculant les pourcentages en fonction des effectifs totaux des cinq.
Observez les résultats du total. Ils sont toujours au « milieu » des résultats des cinq collèges. Par exemple pour les réponses « +++ », le total est à 12 % qui est compris entre la plus basse valeur qui est 8% (Clermont) et la plus forte 18 % (Beauvais). On le remarque plus facilement sur les représentations graphiques comme celle-ci-dessous, où le total est en noir continu et les collèges en pointillés gris.



Les données reçues par l'ordinateur sont au nombre de cinq (une par collège). La sixième courbe est créée par l'ordinateur : c'est la distribution de l'ensemble des collèges.

L'observation, dans tous les cas, de la position de la sixième courbe « entre » celles de chaque collège amène le scientifique à se demander si c'est un heureux hasard ou une vérité absolue.
Par le raisonnement mathématique l'auteur a **démontré** que ce résultat est général.
Le mot **démontré** signifie que grâce au raisonnement on est capable de dire que la courbe « TOTAL » sera « entre » les autres courbes dans toutes les situations pratiques rencontrées et ceci avant même de saisir les données sur l'ordinateur !!

⇒ Seconde situation

On constate dans cette enquête que les courbes sont proches les unes des autres dans de nombreux cas (95 questions sur 123).
Peut-on démontrer qu'il y a toujours un nombre important de questions où les courbes des graphiques correspondants sont proches ?
La réponse est non.
Pour que cela soit vrai, il faudrait définir les mots. Que veut dire un « nombre important » ? 90 % ? 50 % ? 10 % ?
Mais pour que cela soit faux il suffit de trouver un contre-exemple, c'est-à-dire un questionnaire où 100% des réponses donnent des graphiques où toutes les courbes sont éloignées. C'est facile puisque si 95 questions sur 123 correspondent à l'observation souhaitée, il suffit de ne garder dans un questionnaire que les 28 autres questions pour lesquelles des courbes des graphiques ne sont pas proches.
Ces deux situations ont montré que la mathématisation des données permet de clarifier la pensée entre ce qui est vrai et ce qui ne l'est pas. Mais l'approche mathématique de la compréhension des données permet d'aller plus loin. Le mathématicien quand il observe que pour un nombre important de questions les courbes sont proches se demande comment il peut utiliser cette observation qui n'est pas générale mais fréquente ?
C'est ce qu'a fait l'auteur.
Précisons dans un premier temps que la notion de proximité de courbes est une notion bien définie et très connue des mathématiciens. Ceci étant dit l'auteur a raisonné de la façon suivante : « Si on ne peut pas affirmer que les courbes sont proches on peut faire l'hypothèse que l'on ne sélectionne que les questions pour lesquelles c'est le cas. Alors peut-on en déduire sur cette catégorie de questions des propriétés particulières ? »
Alors que dans le cas général la réponse est non, dans ce cas particulier, la réponse est oui.
Pour aller plus avant dans cette réflexion, il faut utiliser le langage mathématique qui, vous l'aurez compris, n'est pas le langage de la vérité mais celui de la clarté des idées.

Jules le 20 janvier 2020