

Effet de structure: de quoi parle t-on... et sans passer par la régression

M.Thevenin

21/04/2022

Exemple des taux d'admission à Berkeley

- https://en.wikipedia.org/wiki/Simpson%27s_paradox
- Données sur les 6 principaux départements de l'université
- Construction d'une base de données individuels sur les principaux indicateurs: petites différences en raison des arrondis (cf base berkeley.csv)

Taux moyen d'admission entre garçon et filles dans l'ensemble des départements

Admission	Sexe		Total
	Homme	Femme	
1	45.99	30.22	39.46

Taux moyen d'admission entre garçon et filles dans l'ensemble par département

	Sexe		Total
	Homme	Femme	
Departement			
A	.6201456	.8301887	.644086
B	.6308244	.6956522	.6333907
C	.3684211	.3401015	.3501094
D	.3301205	.3485255	.3388325
E	.2751323	.2378517	.25
F	.0555556	.0678466	.0623974

Pour mémoire les effectifs:

Departemen t	Sexe		Total
	Homme	Femme	
A	824	106	930
B	558	23	581
C	323	591	914
D	415	373	788
E	189	391	580
F	270	339	609
Total	2,579	1,823	4,402

Taux de candidatures féminines par département

Departemen t	Sexe		Total
	Homme	Femme	
A	88.60	11.40	100.00
B	96.04	3.96	100.00
C	35.34	64.66	100.00
D	52.66	47.34	100.00
E	32.59	67.41	100.00
F	44.33	55.67	100.00
Total	58.59	41.41	100.00

- Le taux moyen de candidatures des filles est en moyenne de 41%, mais varie énormément d'un département à l'autre.
- Le taux moyen d'admission des filles est largement inférieur à celui des garçons, alors que ce n'est jamais le cas lorsqu'on regarde département par département
- Quel serait cet écart moyen des taux d'admission si on neutralisait les fortes différences observées dans les taux de candidatures
- Sans passer par la régression:
 - On peut repondere les observations de sorte que dans chaque département le taux de candidatures féminines soit égal au taux moyen soit 41%.
 - Pour les filles le poids est toujours égal à 1.
 - Pour les garçons il sera inférieur à 1 si dans le département le taux de candidatures féminine est inférieur à 41%, et il sera supérieur à 1 si le taux est supérieur à 41%.

Poids accordés aux garçons dans chaque département

dep	Mean
A	.1819882
B	.0583121
C	2.588509
D	1.271526
E	2.92671
F	1.776236
Total	1

Calcul pour A: Il y a 1.41 fois plus de garçons que de filles tous départements confondus, et 7.78 fois plus dans le département à. Donc pour retrouver le taux moyen, le poids pour les garçons est égal à: $poids_{garçons,A} = \frac{1.41}{7.78} = 0.182$.

Pour tous les départements la structure garçon/fille des candidatures est bien identique à celle calculée sur l'ensemble des 6 départements?

Departemen t	Sexe		Total
	Homme	Femme	
A	58.59	41.41	100.00
B	58.59	41.41	100.00
C	58.59	41.41	100.00
D	58.59	41.41	100.00
E	58.59	41.41	100.00
F	58.59	41.41	100.00
Total	58.59	41.41	100.00

Si on applique cette pondération aux taux de réussite moyens, on peut réellement les comparer sur la base d'une structure garçon/fille identique dans tous les départements.

Admission	Sexe		Total
	Homme	Femme	
0	69.97	69.78	69.89
1	30.03	30.22	30.11
Total	100.00	100.00	100.00

Il n'y a donc plus d'écarts entre garçons et filles. Le taux d'admission pour les filles reste bien identique (car poids toujours égal à 1) mais celui des hommes passe de 45% à 30%.

Pour les garçons ce nouveau taux d'admission permet de mesurer correctement (et ici causalement) l'écart avec celui des filles. Et il est au final inexistant.