

Méthodologie : La corrélation à deux variables

Le principe : il existe une corrélation entre deux variables si il y a un lien entre elles.

On va mettre en évidence une variable explicative sur laquelle on peut trouver facilement des informations (par exemple, l'âge) et une variable expliquée liée (par exemple, la consommation de bonbons). On va ensuite chercher à savoir si ces deux variables évoluent ensemble (par exemple, la consommation de bonbons varie-t-elle avec l'âge).

Le coefficient de corrélation : il permet de vérifier si deux variables évoluent ensemble.

Il se calcule de la manière suivante :

$$r = \frac{\sum X_i Y_i}{\sqrt{\sum X_i^2} \times \sqrt{\sum Y_i^2}} \quad \text{avec } -1 < r < +1$$

où :

- les $X_i = x_i - \bar{x}$ Les x_i représentant les valeurs prises par la variable explicative

- les $Y_i = y_i - \bar{y}$ Les y_i représentant les valeurs prises par la variable expliquée.

Plus r est proche de -1 ou de $+1$, plus la corrélation est forte et plus les deux variables sont liées

On considère que le lien de corrélation est significatif lorsque $r > 0,85$ ou $r < -0,85$.

Si r est positif, les deux variables évoluent dans le même sens.

Si r est négatif, les deux variables évoluent en sens inverse.

Exemple : La société Roby commercialise des bonbons. Elle souhaite vérifier s'il existe une corrélation entre l'âge des consommateurs (x) et la quantité de bonbons (y) en kg consommée durant une année.

Le tableau suivant permet de calculer le coefficient de corrélation :

	x_i	y_i	X_i	Y_i	$X_i Y_i$	X_i^2	Y_i^2
	15	15	-22,5	6,25	-140,625	506,25	39,0625
	30	13	-7,5	4,25	-31,875	56,25	18,0625
	45	5	7,5	-3,75	-28,125	56,25	14,0625
	60	2	22,5	-6,75	-151,875	506,25	45,5625
Total	150	35	0	0	-352,5	1125	116,75
Moyennes	37,5	8,75					

$$r = -0,973$$

$$r = \frac{\sum X_i Y_i}{\sqrt{\sum X_i^2} \times \sqrt{\sum Y_i^2}}$$

Interprétation : dans ce cas, on voit que r est très proche de -1 , on peut donc dire que plus l'âge des consommateurs augmente, moins ils consomment de bonbons.

On met donc en évidence les segments les plus intéressants pour l'entreprise : les jeunes consommateurs (et on pourra même faire des prévisions de consommation de bonbons en fonction de l'âge des consommateurs ciblés).