

C'EST BON A SAVOIR!

Tester une corrélation sans table ...

Donnez-moi un nombre entre -1 et 1? Prenons 0,18. C'est la corrélation.

Donnez-moi un entier entre 10 et 10 000? Prenons 900. C'est la taille n des échantillons : 900 valeurs "x" et 900 valeurs "y".

Je peux vous dire si cette corrélation est significativement différente de zéro (au seuil approximatif 5% par exemple).

On sait que⁽¹⁾, sous l'hypothèse nulle d'une corrélation égale à zéro, la distribution du coefficient a une espérance nulle — ce qui est bien clair — et une variance égale à $1 / (n-1)$ quelles que soient les distributions (inconnues d'ailleurs) de X et Y.

Autour de son espérance 0, la distribution (inconnue, symétrique) du coefficient de corrélation pourra toujours être approchée par la loi normale. La corrélation est donc significative si la valeur est au delà de 2 écarts-types. La question est donc : est-ce que 0,18 est supérieur à 2 sur racine de $(900-1)$ ou encore 2 sur racine de 900? Est-ce que 0,18 est supérieur à 1 sur 15? Tout le monde saura répondre!

Alain Morineau
CISIA•CERESTA

⁽¹⁾ Ce théorème, pourtant très puissant, semble être assez largement ignoré. Rappelons-le : quelles que soient les distributions⁽²⁾ de X et Y, la distribution du coefficient de corrélation sous l'hypothèse d'indépendance a une variance égale à $1 / (n-1)$. La démonstration est simple si X et Y sont des variables discrètes, mais le résultat est vrai dans tous les cas.

⁽²⁾ Sous réserve d'un minimum de bonne tenue mathématique....