

Troisième Colloque national "Contrôle de gestion hospitalier"

CORRELATION N'EST PAS CAUSALITE





Nathalie L'Hostis, Contrôleuse de gestion hospitalière Senior
2 avril 2026

Déclaration d'absence de conflit d'intérêt

En dehors de la société Pilar que j'ai fondée et que je dirige, je n'ai eu au cours des trois dernières années aucune affiliation, intérêt financier ou intérêt de tout ordre avec une société commerciale et ne reçois aucune rémunération, redevance ou octroi de service d'une société commerciale.

Nathalie NARP - L'HOSTIS
2 avril 2026

1. Rappels mathématiques

Un classique des calculs statistiques



*Le **coefficient de corrélation** est un indicateur statistique permettant de mesurer la force et le sens de la liaison linéaire entre deux variables quantitatives.*

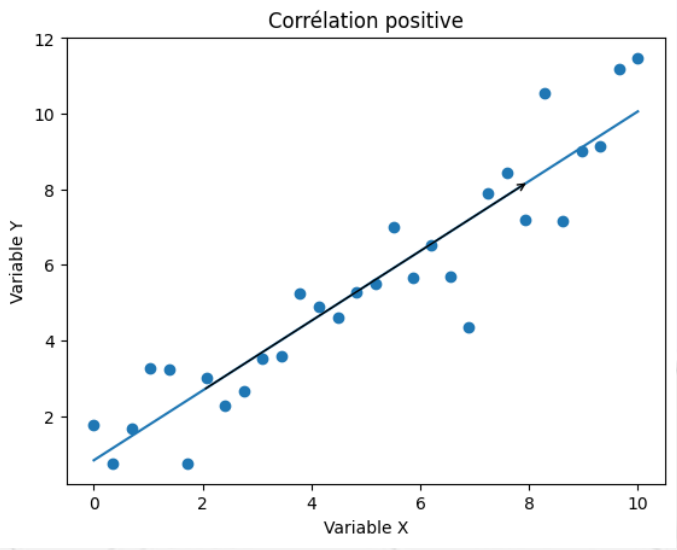
Noté *r*, il est compris **entre - 1 et 1**.

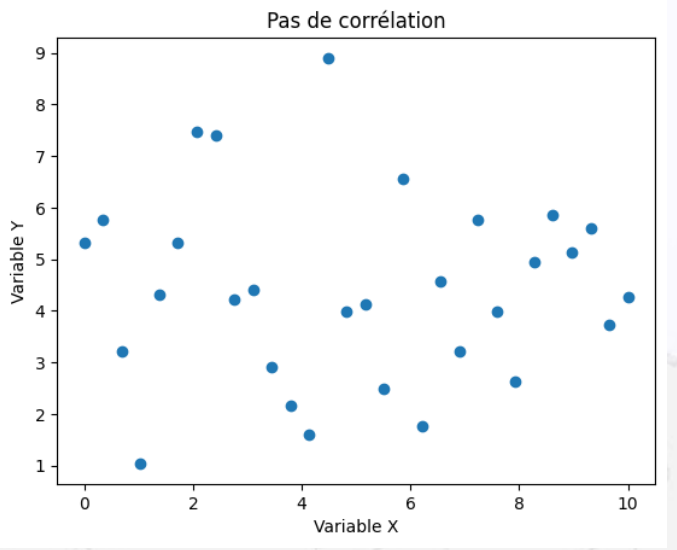
Conséquence :
Les deux séries de données doivent être quantitatives (numériques)

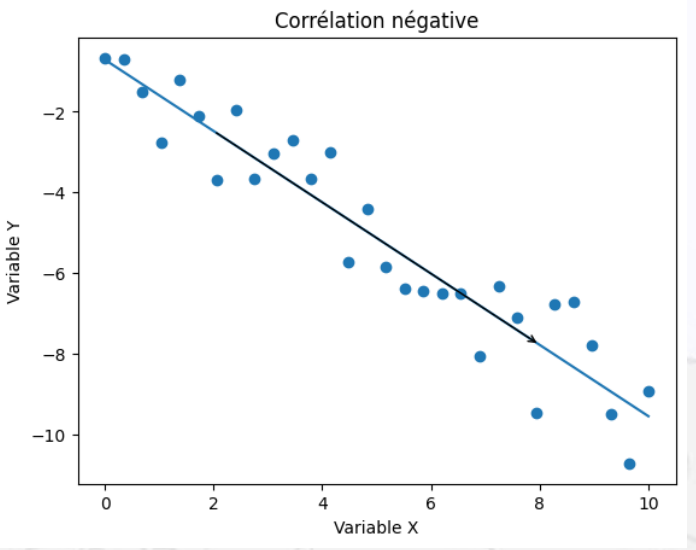


*Le **coefficient de corrélation** est un indicateur statistique permettant de mesurer la force et le sens de la liaison linéaire entre deux variables quantitatives.*

Noté *r*, il est compris **entre - 1 et 1**.
Si sa valeur absolue est supérieure à 0,80, il y a une liaison forte entre les deux séries (selon le contexte)
Si la valeur est positive, les séries évoluent dans le même sens
Si elle est négative, elles évoluent en sens opposé.







*Le **coefficient de corrélation** est un indicateur statistique permettant de mesurer la force et le sens de la liaison linéaire entre deux variables quantitatives.*

La formule compare, pour deux variables, les **écarts de chaque valeur à la moyenne** et les utilise pour indiquer dans quelle mesure la relation entre les variables peut être décrite par une relation linéaire.

$$r = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}}$$

Il existe plusieurs types.
Le plus utilisé est le coefficient de **Pearson** notamment en analyse statistique standard, en économie et en contrôle de gestion, dans les logiciels usuels (Excel, R, Python par défaut)

Pourquoi utiliser le coefficient de corrélation ?

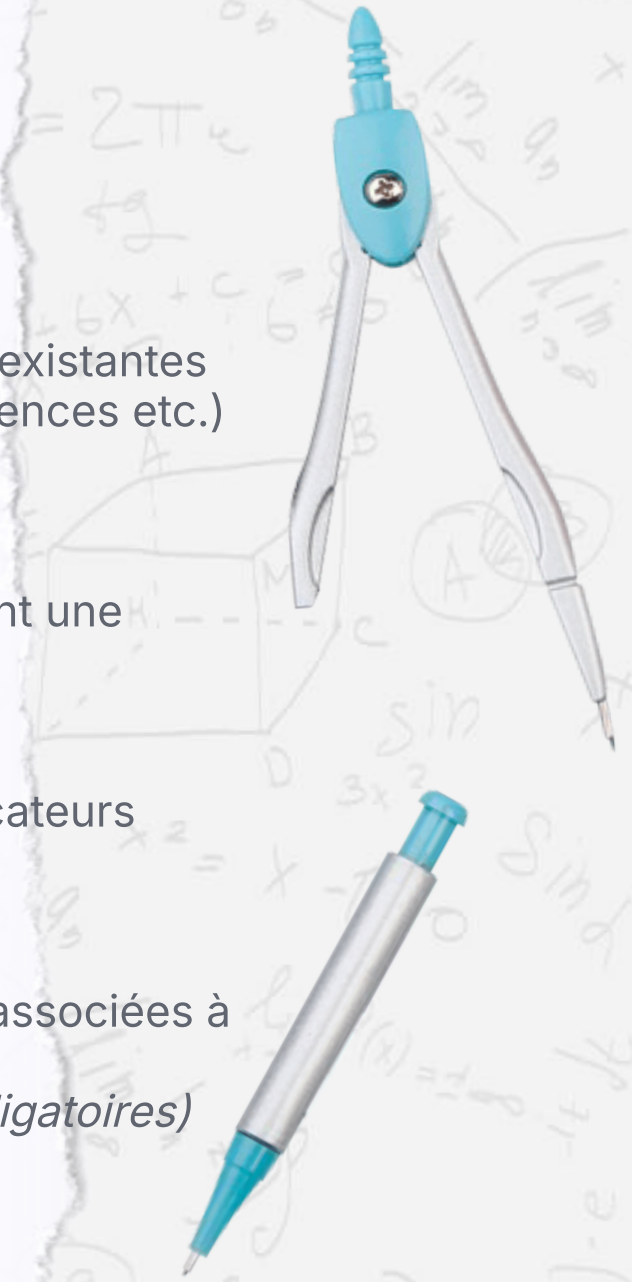
Le coefficient de corrélation s'utilise de façon ponctuelle, en appui d'analyses existantes et de résultats déjà établis (absentéisme, taux d'occupation, passages aux urgences etc.)

Il peut permettre :

- d'approfondir les résultats d'une **analyse médico-économique** en identifiant une relation entre activité et coûts.
  Exemple : nombre de séjours (PMSI) et achats médicaux d'un service

- d'appuyer l'**interprétation d'indicateurs existants** en vérifiant si deux indicateurs évoluent conjointement.
  Exemple : durée moyenne de séjour (DMS) et taux d'occupation des lits.

- d'identifier des **leviers de pilotage** opérationnel en repérant des variables associées à la performance ou aux tensions.
  Exemple : nombre d'accidents du travail et taux de suivi des formations obligatoires)



Des utilisations ciblées



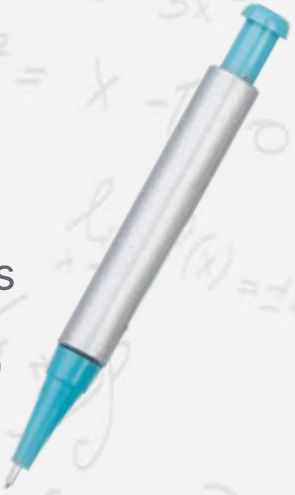
Le coefficient de corrélation ne s'utilise pas de manière récurrente et au cas par cas. Il vient en appui de résultats déjà établis (absentéisme, taux d'occupation, passages aux urgences etc.)

Il peut permettre :

- d'approfondir les résultats d'une **analyse médico-économique** en identifiant une relation entre activité et coûts.
  Exemple : nombre de séjours (PMSI) et achats médicaux d'un service

- d'appuyer l'**interprétation d'indicateurs existants** en vérifiant si deux indicateurs évoluent conjointement.
  Exemple : durée moyenne de séjour (DMS) et taux d'occupation des lits.

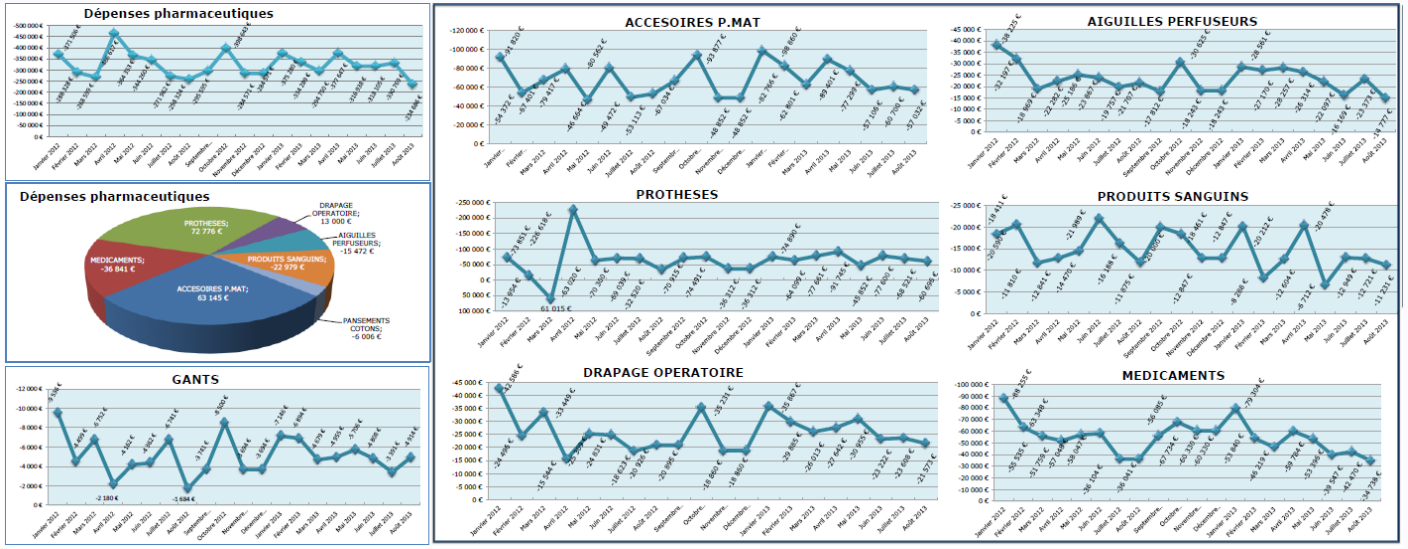
- d'identifier des **leviers de pilotage** opérationnel en repérant des variables associées à la performance ou aux tensions.
  Exemple : nombre d'accidents du travail et taux de suivi des formations obligatoires)



Pourquoi utiliser le coefficient de corrélation ?

Un **exemple** issu d'un tableau de bord de la pharmacie.

Calcul du coefficient de corrélation entre le **nombre de séjours** par mois et les **dépenses pharmaceutiques** par mois.
Les chiffres ont été modifiés par rapport aux originaux



Coefficient de corrélation

Le coefficient de corrélation définit le degré de dépendance entre deux variables. C'est un nombre compris entre -1 (évolution strictement inverse des deux variables) et +1 (évolution strictement parallèle).

Plus il s'éloigne de zéro, plus forte est la corrélation.

r = + 1
r = - 1
r = 0

Ci-dessous, le calcul du coefficient de corrélation entre le nombre de séjours par mois et les dépenses pharmaceutiques par mois.

Dépenses pharmaceutiques	**-0,15**
ACCESOIRES P.MAT	-0,39
PROTHESES	-0,35
MEDICAMENTS	-0,24
DRAPAGE OPERATOIRE	-0,39
AIGUILLES PERFUSEURS	-0,22
PRODUITS SANGUINS	-0,14

2. Pièges et mésusages

Corrélé n'est pas causé



Limites et précautions



Absence de causalité

Deux variables peuvent évoluer ensemble sans lien direct.
Exemple : nombre de séjours et dépenses pharmaceutiques (possible effet saisonnier ou épidémiologique).

Sensibilité aux valeurs extrêmes

Quelques valeurs atypiques peuvent modifier le résultat.
Exemple : un évènement exceptionnel (plan blanc) augmente fortement activité et dépenses.

Limitation aux relations linéaires

Une relation peut exister sans être détectée si elle n'est pas linéaire.
Exemple : relation non proportionnelle entre âge et durée de séjour .

Variable cachée (explication commune)

Une troisième variable peut expliquer les deux autres.
Exemple : la gravité des patients augmente la durée de séjour et les coûts.

Qualité des données

Des données imprécises ou hétérogènes biaisent le résultat.
Exemple : différences de codage PMSI entre services.

Le mésusage majeur :



CORRELATION

VS



CAUSALITE

⚡ **Ne pas interpréter une corrélation comme une relation de cause à effet** ⚡

3. Des sites web décalés

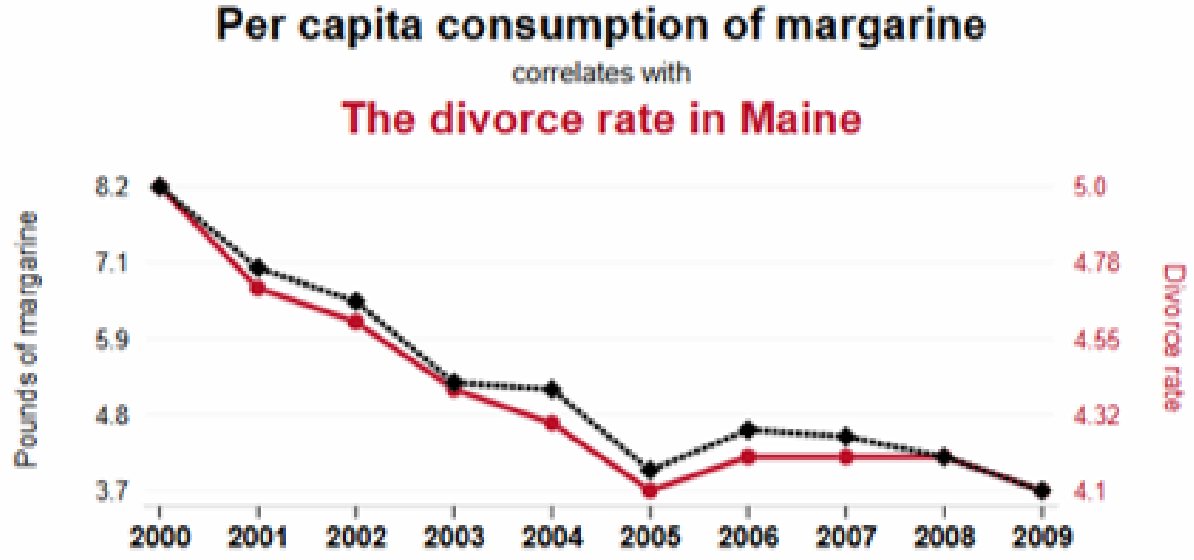
L'humour pour mieux comprendre les biais



https://tylervigen.com/

Spurious Correlations

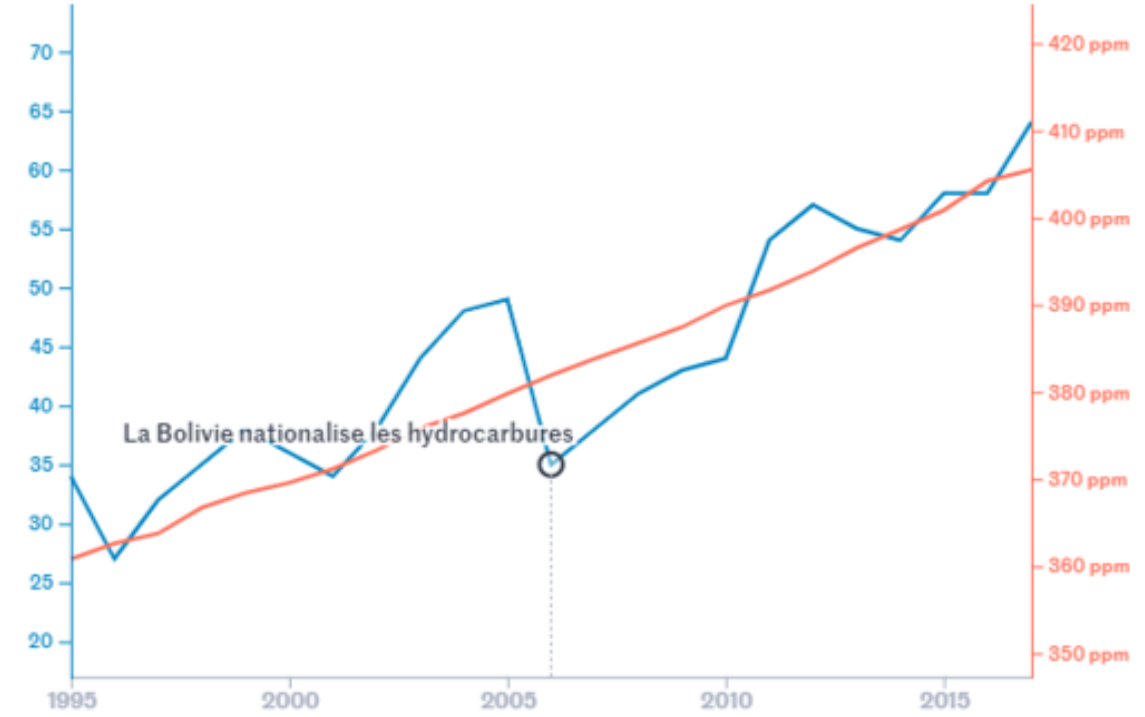
- **Originally published:** May 2014
- **Update launched:** January 27, 2024
- **What it is:** Random correlations dredged up from silly data, turned into linear line charts that showcase the dangers of correlating without cause.
- **More about spurious correlations**



https://www.lemonde.fr/les-decodeurs/article/2019/01/02/correlation-ou-causalite-brillez-en-societe-avec-notre-generateur-aleatoire-de-comparaisons-absurdes_5404286_4355770.html

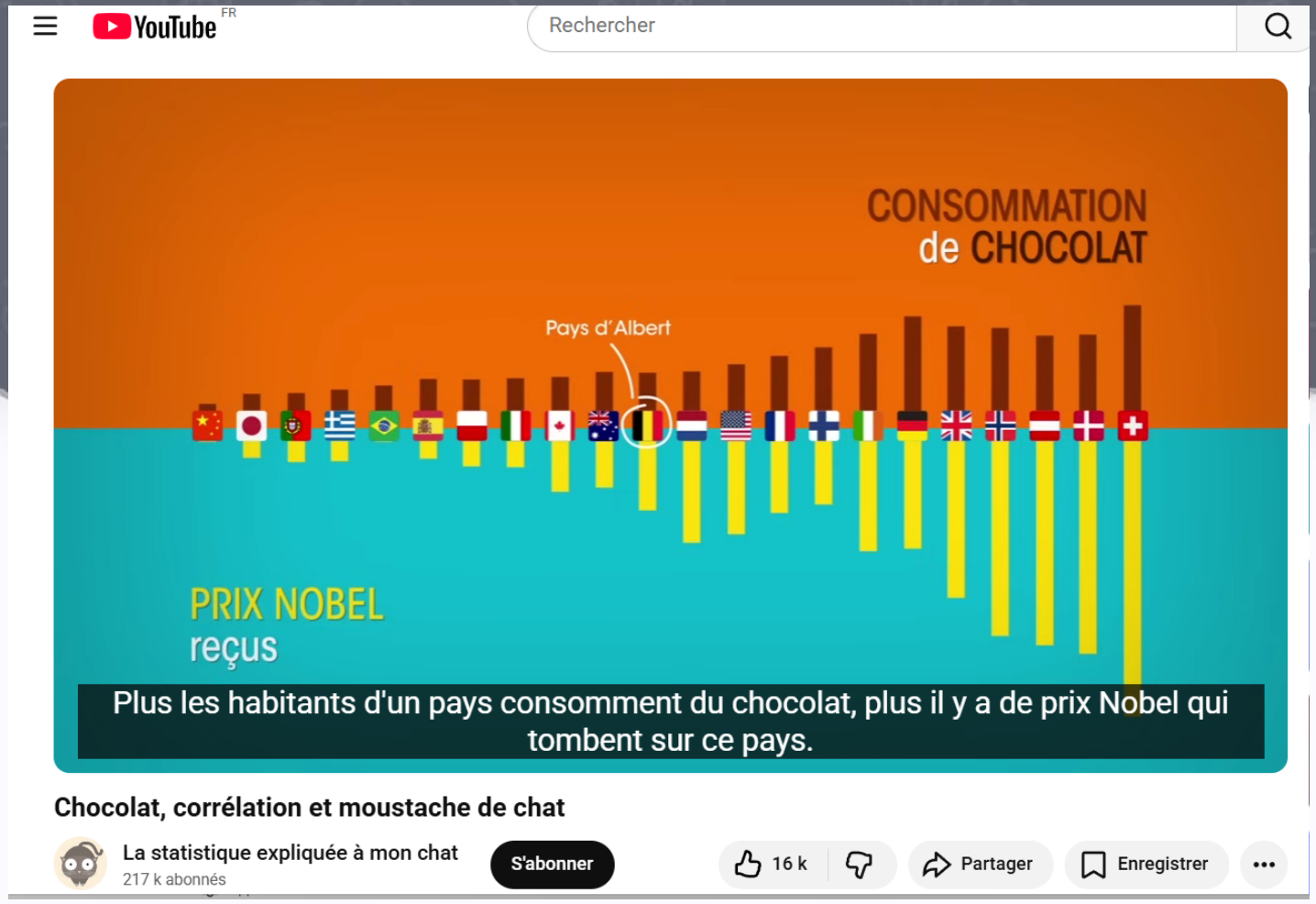
Ce graphique présente de manière aléatoire deux indicateurs – à partir d'environ 50 jeux de données historiques – selon que leur représentation se ressemblent, ainsi qu'un événement au hasard. Ses axes des ordonnées sont coupés en bas pour faire correspondre les courbes.

Cliquez sur le bouton pour générer une nouvelle comparaison au hasard : ici, nous avons un coefficient de corrélation de 90,1 % entre les deux courbes.



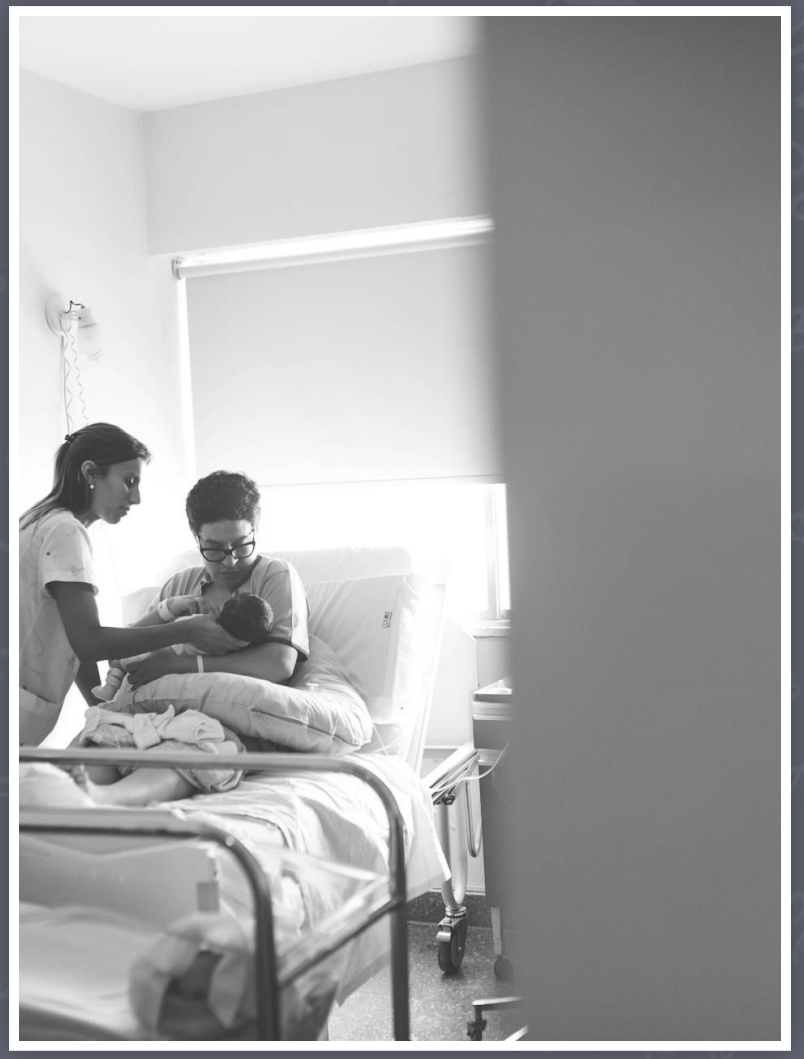
Nouveau-nés français prénommés Lucien
Concentration du CO_2 dans l'atmosphère (en ppm)

https://www.youtube.com/watch?v=aOX0pIwBCvw



4. Illustrations hospitalières

Des cas précis pour enrichir les analyses



Les nouveaux arrêts de travail sont-ils plus nombreux le lundi ?

▼ fonction JOURSEM

Date	Nombre de nouveaux arrêts de travail	Jour de la semaine
01/01/2026	1	4
02/01/2026	3	5
03/01/2026	5	6
04/01/2026	0	7
05/01/2026	0	1
06/01/2026	0	2
07/01/2026	1	3
08/01/2026	3	4
09/01/2026	3	5
10/01/2026	5	6
11/01/2026	2	7
12/01/2026	0	1
13/01/2026	5	2
14/01/2026	3	3
15/01/2026	4	4
16/01/2026	0	5
17/01/2026	1	6
18/01/2026	1	7
19/01/2026	2	1
20/01/2026	3	2
21/01/2026	5	3
22/01/2026	5	4
23/01/2026	2	5
24/01/2026	3	6
25/01/2026	0	7
26/01/2026	2	1
27/01/2026	2	2
28/01/2026	2	3
29/01/2026	2	4
30/01/2026	0	5
31/01/2026	1	6

- 0,02362

=COEFFICIENT.CORRELATION(C8:C38;D8:D38)

COEFFICIENT.CORRELATION(matrice1; matrice2)

Les salaires de direction sont-ils corrélés au genre ?

Matricule	Homme (1) Femme (2)	Salaire annuel brut chargé
5411	1	50 000 €
5332	1	70 000 €
9669	1	80 000 €
2819	1	70 000 €
9082	1	90 000 €
9916	1	85 000 €
4338	1	70 000 €
9295	1	75 000 €
3526	1	100 000 €
1515	2	100 000 €
4746	2	110 000 €
7387	2	45 000 €
677	2	50 000 €
4086	2	35 000 €
9719	2	80 000 €
1427	2	75 000 €
689	2	80 000 €
3536	2	90 000 €
3439	2	70 000 €
8912	2	60 000 €
2674	2	50 000 €
6142	2	50 000 €
9637	2	50 000 €
6742	2	75 000 €
350	2	60 000 €
9336	2	55 000 €
8988	2	80 000 €
3856	2	40 000 €
9303	2	60 000 €
1000	2	70 000 €
6519	2	80 000 €

- 0,25130

=COEFFICIENT.CORRELATION(H8:H38;I8:I38)

COEFFICIENT.CORRELATION(matrice1; matrice2)

5. Et la régression linéaire ?

Pour aller plus loin dans les statistiques et les concepts en IA



La régression, un complément pour comprendre le lien avec la corrélation

La régression linéaire va plus loin que la corrélation : elle permet de traduire la relation entre deux variables sous forme d'une équation.
Elle permet de quantifier une relation : passer d'un constat (« variables liées ») à une estimation (« de combien ? »).

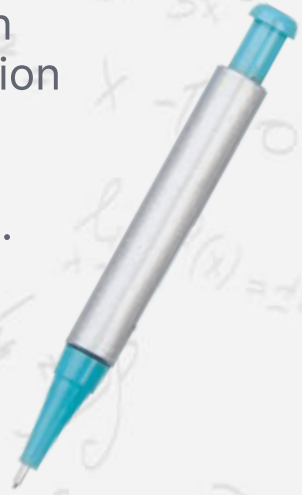
Le coefficient de corrélation indique s'il existe un lien et son intensité, tandis que la régression permet de quantifier ce lien (de combien une variable évolue quand l'autre varie).

Elle permet de tester des hypothèses de gestion (ex : lien entre activité et dépenses, en intégrant plusieurs facteurs) et de mesurer la part des variations expliquées par la relation entre les variables.

Elle repose sur l'**hypothèse d'une relation linéaire**, comme le coefficient de corrélation.

Elle est à la base des outils d'analyse avancée et de prévision, utilisés en IA.





Relation linéaire estimée

Modèle de la forme : $Y = aX + b$

MERCI !

www.pilar-institute.com
33 (0)1 73 02 34 30
contact@pilar-institute.com