

HA1G03X Calling Bullshit

UE Culture Générale

Elodie Brunel

Université de Montpellier – Faculté des Sciences

2024-2025



<https://www.callingbullshit.org/>

"Le monde est submergé par la ... (Bullshit)"

Contre-vérités, fake news, dramatisation... les contenus déformés ou délibérément fallacieux envahissent quotidiennement nos flux. L'opération Calling Bullshit vise à sensibiliser les étudiants à la pensée critique.

- ▶ Janvier 2017 : deux professeurs de l'Université de Washington Jevin West et Carl Bergstrom lancent la plateforme "Calling Bullshit"

« De toutes les choses que j'enseigne à l'université, j'estime que la pensée critique est la compétence la plus essentielle que je puisse inculquer aux étudiants. » , explique Jevin West.



Lire la présentation (en français) de la plateforme dans l'[ADN](#)

Organisation de l'UE

- ▶ Séance 1 : séance introductive
 - Premiers exemples
 - La zététique ? quézaco ?
 - Du bon usage des graphiques
- ▶ Séance 2 : Corrélation Causalité
- ▶ Séance 3 : Discussion autour des "case-studies" de Calling Bullshit et choix de thèmes
- ▶ Séance 4 : Travail encadré sur un thème choisi
- ▶ Séance 5 : Présentation orale du travail en binôme

Evaluation

- ▶ Projet en groupe (2 ou 3 étudiant.e.s)
- ▶ Réalisation d'un dossier à partir d'un cas du site "Calling Bullshit" ou tout autre exemple de votre choix
- ▶ Présentation en classe ou réalisation d'une vidéo de présentation.

Premiers exemples

Exemple 1 : accidents de la route

Les assurances ont établi que plus de 50 % des accidents arrivaient sur un trajet de moins de 30 km. On en a conclu que l'habitude des courts trajets pour aller travailler favorisait le manque d'attention des conducteurs.

Exemple 1 : accidents de la route

Les assurances ont établi que plus de 50 % des accidents arrivaient sur un trajet de moins de 30 km. On en a conclu que l'habitude des courts trajets pour aller travailler favorisait le manque d'attention des conducteurs.

Il est possible que ce soit vrai, mais la « démonstration » est fautive : la plupart des trajets font moins de 30 km !

	≤ 30 km	> 30 km	
nombre de trajets sans accident	990	192	1182
nombre de trajets avec accident	10	8	18
	1000	200	

Exemple 2 : sport et santé

Le conseil de l'Ordre des médecins a publié une étude prouvant que ceux qui pratiquaient régulièrement le jogging à l'âge de 60 ans avaient une probabilité de se trouver en bonne santé à l'âge de 70 ans plus grande que la population normale.

Conclusion de l'Ordre : le jogging est une bonne pratique.

Exemple 2 : sport et santé

Le conseil de l'Ordre des médecins a publié une étude prouvant que ceux qui pratiquaient régulièrement le jogging à l'âge de 60 ans avaient une probabilité de se trouver en bonne santé à l'âge de 70 ans plus grande que la population normale.

Conclusion de l'Ordre : le jogging est une bonne pratique.

Il est encore possible que ce soit vrai, mais ce n'est pas une «démonstration» : la population qui pratique le jogging à 60 ans concentre ceux qui sont déjà en bonne santé. On a donc seulement prouvé que ceux qui sont en bonne santé à 60 ans ont plus de chance de l'être encore 10 ans plus tard.

Exemple 3 : loi des séries

La fameuse « loi des séries » est souvent invoquée en cas de catastrophes quasi simultanées (accidents d'avions, crimes, etc.), Elle n'existe tout simplement pas !

Expérience : S'il se produit en moyenne 12 crashes aériens par an, dessiner sur un axe 12 dates d'accidents « au hasard » :



Exemple 3 : loi des séries

Génération « au hasard » de 12 dates d'accidents avec une machine (un générateur d'aléa) [Excel : fichier accidents.ods](#)

La croyance en la loi des séries repose sur l'idée que s'il se produit 12 crashes en moyenne par an, alors il s'en produira globalement un par mois.

La grande majorité des gens va dessiner des dates à intervalles quasi-réguliers, presque équirépartis.

Exemple 4 : tests PCR /antigéniques

Louis m'a dit :

«Les tests antigéniques sont beaucoup moins fiables que les tests PCR : ma soeur a fait un test antigénique à la pharmacie et elle était négative, mais je lui ai conseillé d'aller faire un test PCR : ils sont beaucoup plus fiables !»

Mais que veut dire Louis ? qu'en est-il vraiment ?

Exemple 4 : tests PCR / antigéniques

Pour comprendre ce que veut dire Louis, il faut connaître la **sensibilité** et la **spécificité** d'un test :

La **sensibilité** est la proportion de vrais-positifs : parmi les individus infectés par SARS-Cov2, ceux dont le test est positif.

La **spécificité** est la proportion de vrais-négatifs : parmi les individus qui ne sont pas infectés par SARS-Cov2, ceux dont le test est négatif.

Les tests antigéniques les plus performants ont une sensibilité d'au moins 80% et une spécificité de 99%, tandis que les tests PCR ont une sensibilité d'au moins 90% et une spécificité de 99,9%.

alors Louis a raison ! allons un peu plus loin ... avec un peu de probas.

Exemple 4 : tests PCR / antigéniques

Supposons qu'en période d'épidémie, la proportion d'individus infectés est de 1% dans la population et qu'on va tester dans cette population un échantillon de 10 000 individus. Construisons un arbre et déterminons la probabilité d'être réellement infecté si le test (antigénique ou PCR) est négatif (faux-négatifs).

Introduction à la zététique ou l'art du doute

Introduction à la zététique ou l'art du doute

Le néologisme « **zététique** » a été (ré-introduit) par Henri Broch, biophysicien, professeur émérite de l'Université Sophia-Antipolis pour désigner un scepticisme critique face aux phénomènes paranormaux.

Le sens du terme est ensuite élargi pour englober d'autres utilisations de la pratique de doute raisonnable sur d'autres sujets. Zététique vient de l'adjectif grec ζητητικός, zētētikós « qui aime chercher »



vidéo [Henri Broch, l'art du doute](#)

L'effet Pangloss

Dans *Candide* (1759) de Voltaire, le précepteur Pangloss prouvait admirablement qu'il n'y a point d'effet sans cause ...
«Il est démontré, disait-il, que les choses ne peuvent être autrement : car tout étant fait pour une fin, tout est nécessairement pour la meilleure fin. Remarquez bien que les nez ont été faits pour porter des lunettes. Les jambes sont visiblement instituées pour être chaussées, et nous avons des chausses,» etc.

Les hypothèses finalistes permettent de faire passer des croyances comme preuves scientifiques. Ce biais de raisonnement est appelé effet Pangloss :



[Vidéo réalisée par des doctorants de Grenoble](#)



[Vidéo version longue](#)

Problèmes de Fermi

Enrico Fermi est un physicien italien connu pour son habileté à donner de bonnes approximations avec peu de données précises. Pendant l'essai atomique du projet Manhattan, il fit ainsi une estimation rapide de la puissance de l'explosion.

Le problème des **accordeurs de piano** est un problème de Fermi classique :

« Combien y a-t-il d'accordeurs de piano à Chicago ? »



[page Wikipedia](#)

Une estimation de Fermi peut être utile pour vérifier la vraisemblance des résultats obtenus après des calculs complexes.



voir aussi l'exemple du site [Calling Bullshit](#)

[Foodstamp fraud](#)

Problèmes de Fermi

Le problème des **accordeurs de piano**

- ▶ il y a environ 5×10^6 d'habitants à Chicago ;
- ▶ en moyenne, il y a deux personnes par foyer ;
- ▶ en gros, un foyer sur vingt possède un piano qu'il faut accorder régulièrement ;
- ▶ les pianos accordés régulièrement sont accordés à peu près une fois par an ;
- ▶ un accordeur de piano met à peu près deux heures pour accorder un piano, en comptant le temps de déplacement ;
- ▶ un accordeur de piano travaille huit heures par jour, cinq jours par semaine, cinquante semaines par an ;

Evaluer le nombre d'accordeurs de piano ! (le nombre réel est environ de 190). Vous pouvez refaire votre estimation en modifiant certaines approximations !

Les paradoxes mathématiques

Le paradoxe de Simpson est un paradoxe mathématique qui peut conduire à des conclusions erronées. Le statisticien Edward Simpson a formulé ce paradoxe en 1951.



[page Wikipedia](#)



à voir dans la série "Voyage au pays des maths" sur arte :
[Vidéo](#)

Du bon usage des graphiques

Du bon usage des graphiques

Lorsqu'on présente des statistiques, il est important de savoir comment illustrer l'information sous forme de graphique ou diagramme. On parle aujourd'hui de **Data Visualisation**.

Voici quelques règles générales : Un graphique doit :

- ▶ illustrer les faits avec précision,
- ▶ attirer l'attention du lecteur,
- ▶ compléter les arguments du texte,
- ▶ comporter un titre et des étiquettes,
- ▶ être clair et simple,
- ▶ montrer les données sans modifier leur message,
- ▶ montrer clairement toute tendance ou différence dans les données,
- ▶ être exact en ce qui a trait à l'aspect visuel.

Vous pourrez compléter en allant sur le site de :



Statistique Canada

Du bon usage des graphiques

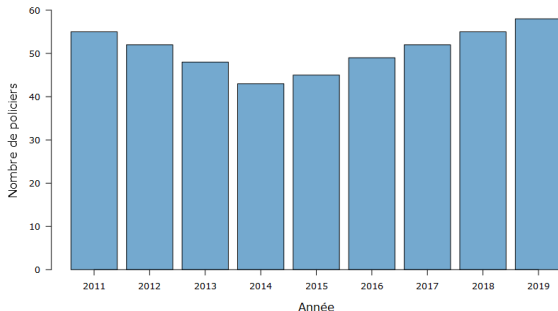
On peut utiliser différents types de graphiques pour diffuser de l'information, notamment :

- ▶ le graphique à barres,
- ▶ le pictogramme,
- ▶ le graphique circulaire,
- ▶ le graphique "temporel",
- ▶ le nuage de points,
- ▶ l'histogramme.

Du bon usage des graphiques

Le graphique à barres

Graphique 5.2.1
Nombre de policiers à Crimeville, 2011 à 2019



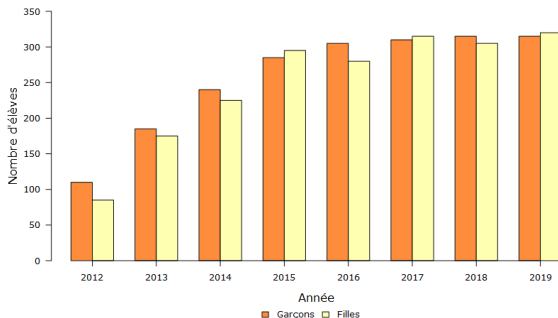
attention les

barres ne doivent pas être "collées" : ce n'est pas un histogramme !

Du bon usage des graphiques

Le graphique à barres

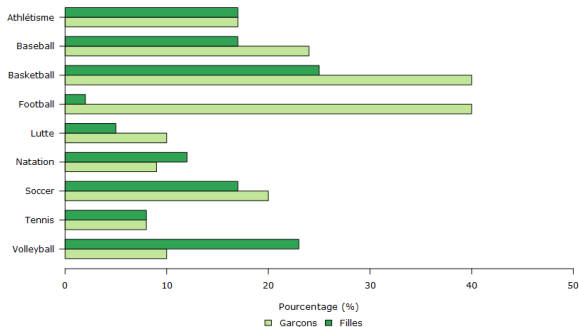
Graphique 5.2.2
Élèves possédant un téléphone intelligent à l'école des Bois-Francis,
selon le genre, 2012 à 2019



Du bon usage des graphiques

Le graphique à barres

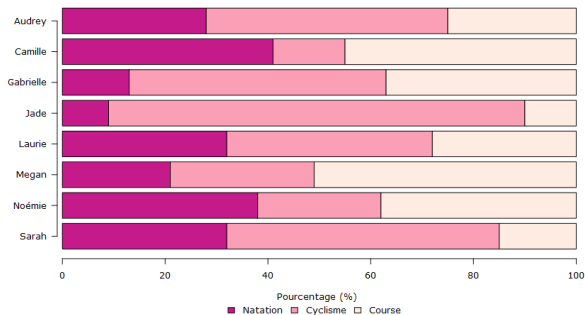
Graphique 5.2.3
Sports pratiqués par les élèves de 15 ans de l'école de William, selon le genre



Du bon usage des graphiques

Le graphique à barres

Graphique 5.2.4
Triathlon à l'école Rousseau, pourcentage du temps consacré à chaque sport
par les compétitrices



Du bon usage des graphiques

Le pictogramme

Graphique 5.3.2
Pouvoir d'achat du dollar canadien, 2000 à 2020

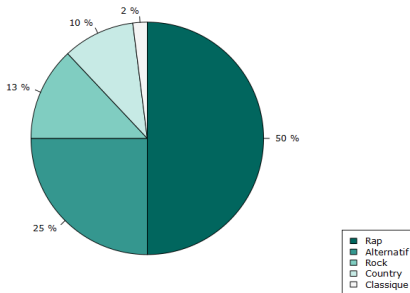


surtout dans le dessin de presse !

Du bon usage des graphiques

Le graphique circulaire Un graphique circulaire est construit en calculant la part de chacune des modalités en un pourcentage de 360 degrés (règle de 3).

Graphique 5.4.2
Genres musicaux préférés des jeunes adultes de 14 à 19 ans

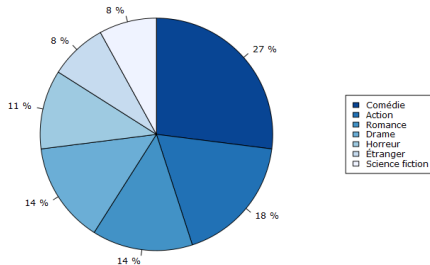


Permet de voir rapidement l'information : pas plus de 4 ou 5 modalités.

Du bon usage des graphiques

Le graphique circulaire

Graphique 5.4.5
Types de film préférés dans la classe de cinéma de Mme Robert

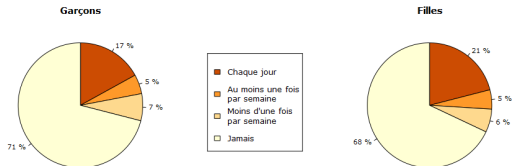


Trop de modalités : préférez un graphique à barres dans ce cas.

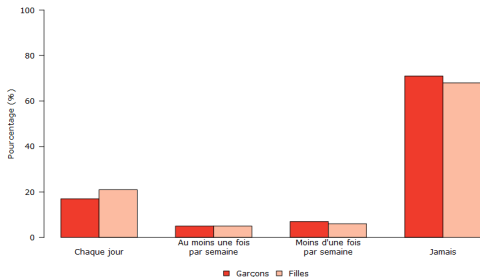
Du bon usage des graphiques

Le graphique circulaire

Graphique 5.4.6
Tabagisme chez les membres de l'équipe d'athlétisme de 15 ans de l'école secondaire du Parc, selon le genre



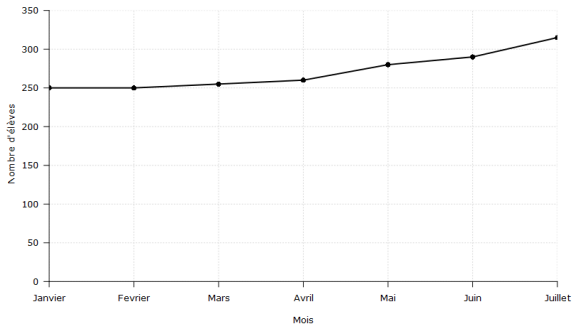
Graphique 5.4.7
Tabagisme chez les membres de l'équipe d'athlétisme de 15 ans de l'école secondaire du Parc, selon le genre



Du bon usage des graphiques

Le graphique "temporel" On représente une variable y sur l'axe des ordonnées en fonction d'une variable x qui représente le temps sur l'axe des abscisses : on relie les points par des segments de droite.

Graphique 5.5.1
Participation à la population active à l'école d'Olivier

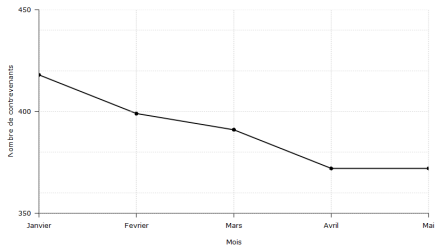


Permet de voir une tendance temporelle.

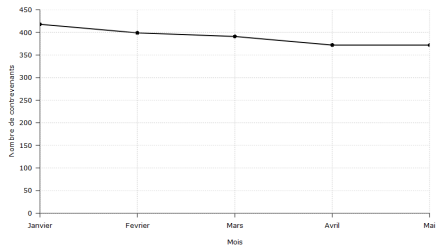
Du bon usage des graphiques

Le graphique "temporel" attention à l'échelle des axes du graphique !

Graphique 5.5.3
Nombre de contrevenants reconnus coupables à Grishamville



Graphique 5.5.4
Nombre de contrevenants reconnus coupables à Grishamville



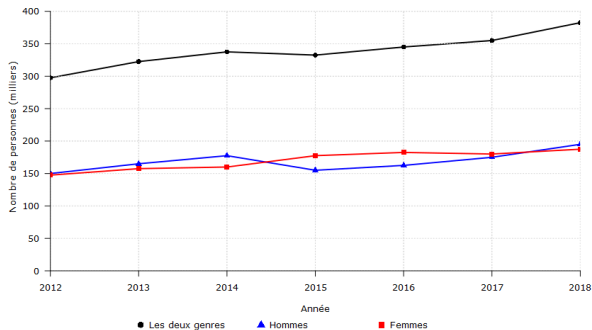
Le 1er graphique exagère la tendance. Le 2nd montre que le déclin des contrevenants est réellement minimale.

Les deux graphiques peuvent avoir leur utilité selon le contexte. La chose importante à retenir, c'est qu'il faut toujours porter attention à l'échelle lors de l'interprétation d'un graphique.

Du bon usage des graphiques

Le graphique "temporel" Ce graphique est un exemple d'un bon graphique. Le message est énoncé clairement dans le titre et chaque ligne du diagramme est identifiée correctement.

Graphique 5.5.5
Utilisation des téléphones cellulaires par genre, Toutedville, 2012 à 2018

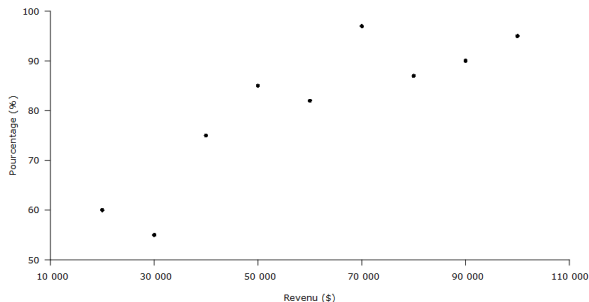


Du bon usage des graphiques

Le nuage de points le nuage de points est grandement utilisé pour présenter la mesure de deux ou plusieurs variables liées. Le nuage de points est particulièrement utile lorsque les valeurs des variables sur l'axe des y dépendent des valeurs de la variable de l'axe des x.

Les points ne sont pas reliés.

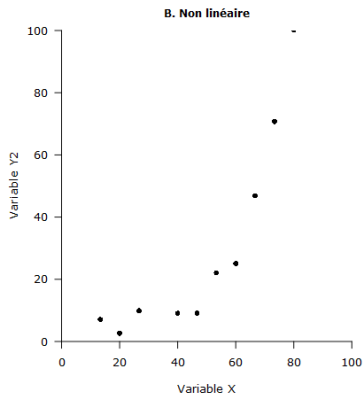
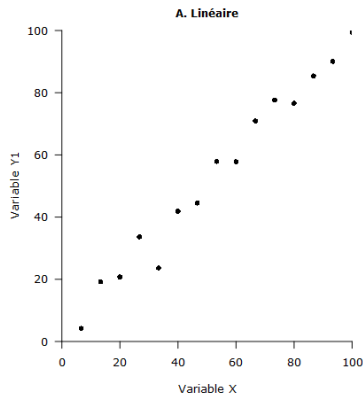
Graphique 5.6.1
Possession d'une voiture à Touteville, selon le revenu du ménage



Du bon usage des graphiques

Le nuage de points Recherche d'un lien linéaire ou non-linéaire entre deux variables.

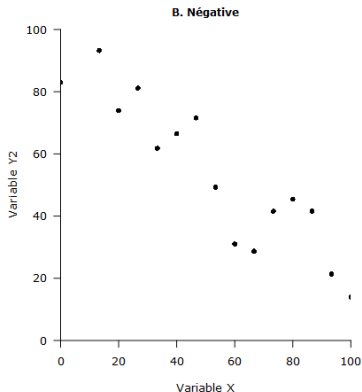
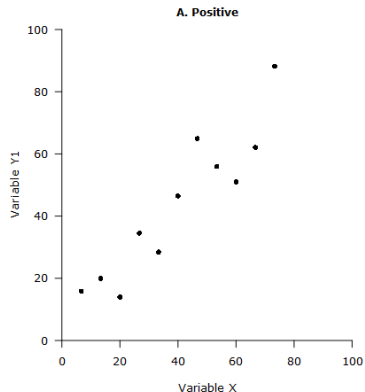
Graphique 5.6.2
Relation linéaire ou relation non linéaire



Du bon usage des graphiques

Le nuage de points Le lien linéaire, quand il existe, peut être positif ou négatif :

Graphique 5.6.3
Relation positive ou relation négative



Du bon usage des graphiques



voir exemples de Calling Bullshit [misleading axes](#)

Corrélation et Causalité

Qu'est-ce que la régression linéaire ?

Régression Linéaire et corrélation

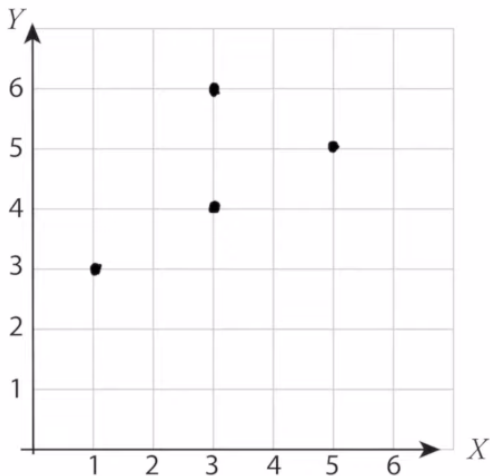
En statistique, un modèle de régression linéaire est un modèle qui cherche à établir une relation linéaire entre une variable Y et une variable X .

Exemple "jouet" : Existe t'il une relation linéaire entre le nombre d'heures passées à étudier les maths X et le résultat au test Y ?

Nombres d'heures X	1	3	3	5
Note au test Y	3	4	6	5

Régression Linéaire et corrélation

Représentation graphique des données : le nuage de points.

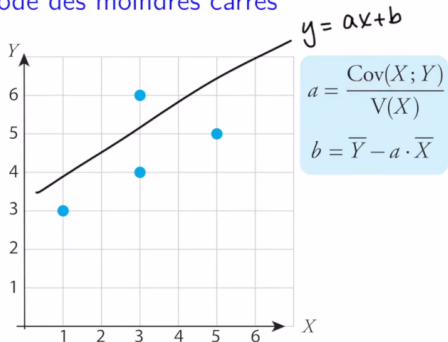


Régression Linéaire et corrélation

On cherche une droite d'équation $y = ax + b$ qui passe "au plus près" des points du nuage. Pour cela on minimise un critère appelé le critère des moindres carrés.

Les valeurs des coefficients a et b qui minimisent ce critère sont :

Méthode des moindres carrés



Régression Linéaire et corrélation

On commence par calculer le coefficient a :

$$a = \frac{\text{Cov}(X; Y)}{V(X)} = \frac{\overline{XY} - \overline{X}\overline{Y}}{\overline{X^2} - \overline{X}^2}$$

Puis on calcule le coefficient $b = \overline{Y} - a\overline{X}$.

Régression Linéaire et corrélation

En pratique, on peut très rapidement faire ces calculs dans un tableur :

	X	Y	XY	X^2	Y^2
	1	3			
	3	4			
	3	6			
	5	5			
Somme	12	18			
Moyenne	3	4,5			

Régression Linéaire et corrélation

On ajoute ensuite les calculs des colonnes produit XY et carré X^2 et Y^2 :

	X	Y	XY	X^2	Y^2
	1	3	3	1	9
	3	4	12	9	16
	3	6	18	9	36
	5	5	25	25	25
Somme	12	18	58	44	86
Moyenne	3	4,5	14,5	11	21,5

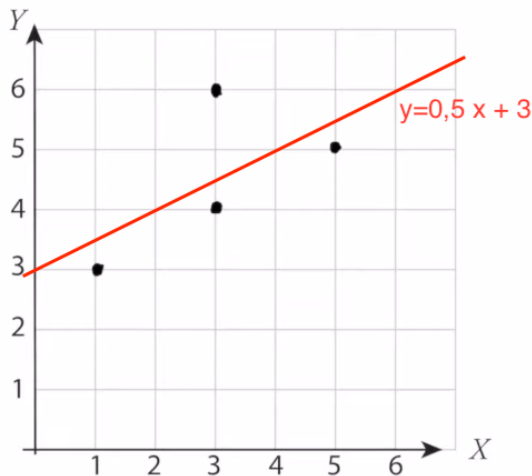
$$a = \frac{\overline{XY} - \bar{X}\bar{Y}}{\overline{X^2} - \bar{X}^2} = \frac{14,5 - 3 \times 4,5}{11 - 3^2} = \frac{14,5 - 13,5}{11 - 9} = 0,5$$

$$b = \bar{Y} - a\bar{X} = 4,5 - 0,5 \times 3 = 3$$

Régression Linéaire et corrélation

On a ainsi obtenu la droite dite de "régression" par la méthode des moindres carrés :

Excel (ou LibreOffice) fichier [regression-jouet.ods](#)



Régression Linéaire et corrélation

Le coefficient de corrélation linéaire (de Pearson)

$$r = \frac{\text{Cov}(X, Y)}{\sqrt{V(X)V(Y)}}$$

mesure la force du lien linéaire entre X et Y .

Propriété : $-1 \leq r \leq 1$.

On déjà calculé $\text{Cov}(X; Y) = 1$ et $V(X) = 2$. Il reste à calculer $V(Y)$:

$$V(Y) = \overline{Y^2} - \overline{Y}^2 = 21,5 - 4,5^2 = 1,25$$

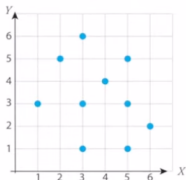
Et donc,

$$r = \frac{1}{\sqrt{2 \times 1,25}} \simeq 0,6$$

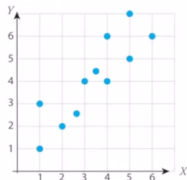
Ici on a une corrélation positive et modérée.

Régression Linéaire et corrélation

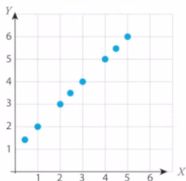
Interprétation du coefficient de corrélation linéaire :



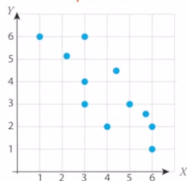
Absence de corrélation



Corrélation positive



Corrélation positive parfaite



Corrélation négative

Corrélation n'est pas toujours Causalité

Une corrélation positive entre deux variables signifie que les deux variables varient dans le même sens (lorsqu'une variable augmente, l'autre augmente également).

Une corrélation est négative, lorsque les deux variables varient en sens inverse. Dans ce cas, lorsqu'une variable augmente, l'autre diminue.

Il y a absence de corrélation lorsque si une variable varie, l'autre ne varie pas, ou alors varie sans lien avec la première.

Différencier causalité et corrélation

Une **corrélation** est un **lien statistique**, sans qu'on se demande quelle variable agit sur l'autre.

Une **causalité** est un lien qui affirme qu'une variable agit sur une autre.

Deux événements peuvent être corrélés, mais ils ne sont pas nécessairement causés l'un par l'autre : une corrélation ne reflète donc pas toujours un lien de causalité entre les deux variables reliées statistiquement.

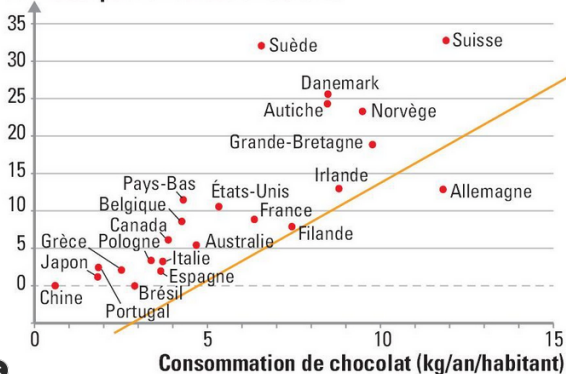
Tout d'abord, une corrélation statistique peut-être :

- ▶ l'effet du hasard.
- ▶ dû à une **variable explicative cachée**, influençant simultanément les deux autres variables.

Exemple

Le scientifique Franz Messerli dans "Chocolate Consumption, Cognitive Function, and Nobel Laureates" (2012) montrait cette célèbre corrélation positive entre la consommation de chocolat par habitant et par an et le nombre de prix Nobel pour 10 millions d'habitants. **Manger du chocolat rendrait-il les individus intelligents ?**

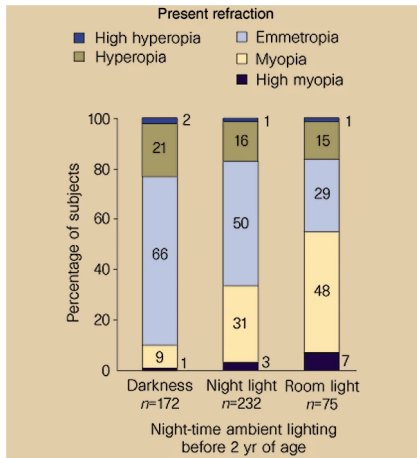
Prix Nobel par 10 millions d'habitants



Exemple

Un article dans la très sérieuse revue [Natures](#) en 1999 démontrait que les enfants soumis à la lumière d'une veilleuse entre 0 et 2 ans avaient davantage tendance à développer une myopie ...

Figure 1: Present refractions of children of ages 2-16 yr and night-time light exposure before the age of 2 yr.



L'effet cigogne



Corrélation n'est pas causalité : les dérives de l'effet cigogne et aussi :



"Chocolat, corrélation et moustaches de chat"



Paradoxe de simpson : "Fumer aide à courir plus vite !"



voir un simulateur de comparaisons absurde ! "Les décodeurs" dans le Monde

Ce qu'il faut retenir ...

Démontrer une théorie avec seulement des statistiques peut être trompeur. Souvent la théorie préexiste et les chiffres sont ensuite utilisés pour la conforter « scientifiquement ».

Quelles hypothèses lorsqu'une corrélation est mise en évidence entre deux variables X et Y ?

- ▶ Y cause X ,
- ▶ X cause Y
- ▶ X et Y ont une cause commune Z ,
- ▶ le fruit du hasard !

Adopter une démarche scientifique ...

Si les corrélations sont faciles à trouver, la relation de cause à effet entre deux situations est beaucoup plus complexe.

Le travail des scientifiques, quel que soit leur domaine d'étude :

Pour évaluer le lien de causalité, il faut établir des **protocoles statistiques**. Ça veut dire :

- ▶ comparer des groupes similaires et représentatifs,
- ▶ avoir un nombre de cas aussi grand que possible,
- ▶ les étudier pendant une durée "suffisamment" longue,
- ▶ pouvoir répéter les expériences.

Une fois l'étude terminée, même si on observe un lien de cause à effet, Il faut alors multiplier les études, et donc les résultats, pour renforcer l'hypothèse. C'est tout le principe de la démarche scientifique.

En plus de tout cela ...

- ▶ qui a produit ces statistiques ?
- ▶ avec quel objectif ?
- ▶ avec quelle objectivité ?
- ▶ selon quelles modalités de calculs ?
- ▶ avec quelles définitions ? (pour faire baisser le taux de chômage, il suffit de changer la définition de ce qu'est un chômeur par exemple).