

Faut-il croire les chiffres ?

Manuel d'autodéfense statistique

Mai 2024



Photo prise par Jean-Pierre Cling à Paris en janvier 2023 lors d'une manifestation contre la réforme des retraites.

Plan

*Chacun de ces chapitres sera illustré par trois ou quatre exemples avec une petite conclusion d'ensemble, chaque exemple étant lui-même assorti d'un rapide commentaire.
Des citations agrémentent le colis.*

Introduction

Quelques notions fondamentales

A) Du général au particulier. Ou du particulier au général.

- A1) Du général au particulier : une histoire probablement pas authentique (on peut l'espérer !)
- A2) L'erreur inverse (passer abusivement du particulier au général) est encore plus répandue.
- A3) Et les vaccins ?

B) Les dangers de la moyenne

- B1) La réunion de millionnaires... et d'un autre
- B2) Variante : Le salaire moyen n'est pas celui qu'on croit
- B3) Attention aussi aux évolutions « moyennes » et aux effets de structure
- B4) Quelles pondérations a-t-on employées ?

C) L'ambiguïté sur l'objet dont on parle (domaine concerné, pratiques évoquées...)

- C1) Je suis d'une taille supérieure à la moyenne !
- C2) « Je représente 30 % des électeurs »
- C3) Il y a beaucoup plus de personnes qui meurent dans un lit... qu'en tombant dans un précipice en montagne.
- C4) Un enfant qui naît aujourd'hui en France est-il gravement endetté ?

D) Les traquenards des sondages d'opinion

- D1) Unanimité factice : « Les Parisiens » pensent que...
- D2) Une réponse « plus fréquente » ne définit pas une majorité
- D3) Un sondage n'est qu'un sondage, même s'il est bien fait
- D4) « Les Parisiens » pensent que... Mais comment leur a-t-on posé les questions ?

E) Les « simples » erreurs de calcul ou les utilisations erronées d'une méthode, d'une notion mathématique inappropriée

- E1) Ne pas confondre points et pourcentages !
- E2) Ne pas confondre corrélation et causalité
- E3) Les chiffres ne parlent jamais d'eux-mêmes.
- E4) Les 80/20 ne sont pas une loi statistique

Conclusion

Bibliographie commentée

Quelques sites et articles

Citations choisies

*On pourra trouver ce manuel ici :
sites.google.com/view/statisticiens-espace-rural/accueil
et www.cgтинsee.org
ou le demander à alaingely@orange.fr*

Introduction

Devant le déferlement de nombres, de tableaux, de graphiques prétendant décrire la réalité, nous sommes parfois tentés :

- de les gober sans esprit critique en se disant confusément « Il y a des chiffres, cela doit donc être scientifiquement établi par des experts » ;
- ou, à l'inverse, de tout rejeter en bloc en affirmant que « Les statistiques sont fausses »... sans esprit critique non plus !
- ou encore de prétendre pouvoir tout quantifier, de vouloir évaluer toutes les activités par rapport à des objectifs chiffrés qui imposeraient des décisions.

Non, un indicateur (statistique) ne doit pas être un dictateur. Comme les panneaux indicateurs, il ne fournit que des... indications, éventuellement erronées. Et il ne saurait nous dicter où nous devons aller.

Il est vrai qu'il n'est pas toujours facile de s'y retrouver. Comment trier :

- le bon grain (les statistiques qui aident à décrire et à comprendre le monde, car elles existent !)
- de l'ivraie (les nombres qui, par incompetence ou par malhonnêteté de la part de leurs diffuseurs, nous fourvoient et obscurcissent notre vision de la réalité bien plutôt que de l'éclairer) ?

Notre point de départ sera une formule non mathématique. Quantifier = convenir + mesurer

Cet énoncé d'Alain Desrosières, historien et sociologue de la statistique, offre une clé d'accès à l'analyse critique de beaucoup de ces nombres qui nous sont présentés.

En effet, quantifier (c'est-à-dire faire naître des nombres) suppose deux choses :

- que le producteur des chiffres et leur utilisateur sachent de quoi ces chiffres parlent. Par exemple « les chiffres du chômage » n'ont de sens que face à des définitions de cette notion, résultant de « **conventions discutées et si possible acceptées** ». Nombreuses sont les personnes qui se considèrent comme chômeuses, ou plus largement « privées d'emploi », mais ne correspondent pas à la définition officielle.
- un mode d'élaboration, une **méthodologie qui permet de « mesurer »** le phénomène considéré. Pour reprendre l'exemple du chômage, les deux principales sources statistiques recourent à des méthodes très différentes. Leurs résultats sont parfois divergents mais complémentaires. D'une part, France Travail est une source administrative très détaillée. Elle quantifie des « demandeurs d'emploi », qui sont le produit des catégories, des normes et des missions assignées à cette administration. D'autre part, le chômage officiel est connu grâce à l'enquête sur l'Emploi de l'Insee. Cette enquête décrit la réalité sociale de l'emploi et du chômage, selon les définitions du bureau international du travail. C'est un sondage sérieux de la statistique publique. Il est en l'occurrence bien documenté, ce qui n'est pas toujours le cas des sondages privés qui envahissent les médias. Mais ses résultats n'ont de sens qu'en référence à des définitions précises, à des nomenclatures.

Un exemple : une personne qui a effectué une seule heure de travail rémunéré au cours de la semaine précédant l'enquête n'est, de ce fait, pas considérée comme chômeuse mais comme « en emploi » selon les définitions du Bureau international du travail, reprises par l'Insee.

<https://www.insee.fr/fr/metadonnees/definition/c1159>

On admettra que c'est là une définition restrictive du chômage !

En résumé, on peut traduire ainsi la formule *Quantifier = convenir + mesurer* en disant qu'un chiffre n'a pas de sens en soi, dans l'analyse économique, le débat social ou dans la controverse scientifique. Proposer un nombre (= quantifier) repose nécessairement sur une définition, un ensemble de conventions (= convenir) et aussi sur des méthodes d'observation rigoureuses et explicites (= mesurer).

Beaucoup de « pièges de la statistique » s'expliquent par des entorses à, au moins, l'une de ces deux exigences (convenir et mesurer). D'autres procèdent d'un défaut de connaissance des ordres de grandeur ou des lois de l'arithmétique, notamment.

Cette page n'est pas indispensable pour lire la suite mais elle peut être utile !

On n'utilisera pas de formules mathématiques dans ce manuel.
Rappelons quand même **quelques notions fondamentales** en les illustrant.

La moyenne (arithmétique) c'est la somme des valeurs divisée par le nombre de ces valeurs.

$(X_1 + X_2 + X_3 + \dots + X_n) / n$

C'est une quantité abstraite qui n'a pas toujours de sens concret

Exemple : deux sœurs ont, respectivement, mis au monde deux et trois enfants.

La moyenne est donc : $(2+3)/2 = 2,5$ enfants. Sans commentaire.

Un pourcentage c'est, par définition, cent fois le rapport correspondant .

Exemples : une liste a obtenu 200 voix, soit V, sur S = 1000 suffrages exprimés.

Le rapport $V/S = 200/1000 = 0,2$. Soit en pourcentage : $(V/S)*100 = 0,2 * 100 = 20 \%$.

Se référer à une population (fictive) de 100 individus, cela peut faciliter des comparaisons par rapport à une élection précédente, ou aux résultats du même parti au niveau national.

Remarque : parler en % peut aussi être un moyen de présenter une réalité de manière tendancieuse

1er exemple : si X a obtenu 2 voix sur 10 votants, lui attribuer 20 % tend à grossir un peu artificiellement son audience.

2ème exemple : s'il y a 5 millions de chômeurs sur une population active de 25 millions, il n'est pas faux de publier un taux de chômage de $(5/25)*100$ soit 20 %, mais on minimise le problème social si on ne signale pas aussi qu'ils sont 5 millions.

L'évolution en % d'une variable P entre deux dates t0 et t1 est le résultat du calcul $100*(P_1-P_0)/P_0$

Exemple fictif : P est le prix de la baguette de pain. Il serait passé de 0,90 à 1 euro entre 2023 et 2024.

L'évolution de ce prix est égale à : $(1-0,90)/0,90 = 0,10/0,90 = 0,111$. En pourcentage : $0,111 * 100 = 11,1 \%$

L'année 0, ici 2023, est parfois dite « année de base », souvent ramenée à un « indice 100 ».

On dirait alors que le prix de la baguette en 2024 « base 100 en 2023 » est égal à 111,1.

Remarque : On peut dire que le prix de la baguette a augmenté de 10 centimes (c'est l'évolution en valeur absolue) ou de 11,1 % (c'est l'évolution en valeur relative). Les deux se justifient mais délivrent souvent un « message » différent.

Attention : l'évolution en valeur relative peut être un peu trompeuse si on part d'un chiffre très faible (dénominateur proche de 0).

l'exemple des téléphones portables quand ils sont apparus et se sont rapidement multipliés. Admettons que le nombre de vols de ces appareils soit passé de 100 la première année à 1000 la seconde année.

Le taux d'accroissement, qui serait ici de $100 * (900/100) = 900 \%$ n'est pas vraiment faux mais franchement fallacieux !

Un tel calcul risque même de n'avoir aucun sens quand la variable peut-être négative, comme par exemple le solde d'un compte

bancaire. Si j'étais dans le rouge la première année de 100 € et rétabli la seconde, avec + 200 €, peut-on parler d'un « accroissement » de - 300 % qui résulterait du calcul : $100 * (200 - (-100)) / (-100)$?

Non ! En revanche, l'évolution en valeur absolue (ici : +300 €) a du sens.

A) Du général au particulier. Ou du particulier au général.

« Les statistiques sont vraies quant à la maladie et fausses quant au malade ; elles sont vraies quant aux populations et fausses quant à l'individu. » Léon Schwartzberg.

A1) Du général au particulier : une histoire probablement pas authentique (on peut l'espérer !)

Tel père d'une famille de neuf enfants refusait absolument un dixième enfant.

Pourquoi ? Parce qu'il avait entendu dire que sur dix enfants qui naissaient, il y avait un chinois. Il croyait que cette statistique s'appliquerait à sa famille. Et il ne voulait pas d'un chinois dans sa maison.

Commentaire sur ce point

Cet exemple est évidemment absurde, en plus d'être raciste. On ne peut appliquer à ce couple le ratio d'un dixième, qui serait valable pour l'ensemble de l'humanité, à une famille et à une date données. Mais ce cas imaginaire traduit une attitude d'esprit, malheureusement fréquente : appliquer abusivement à une situation particulière une donnée statistique.

A2) L'erreur inverse (passer abusivement du particulier au général) est encore plus répandue.

a) Par exemple : incriminer « les Italiens » parce qu'un délit a été attribué à un Italien dans un quartier où vivent des Italiens. Conclusion xénophobe, malheureusement bien réelle, celle-ci : fr.wikipedia.org/wiki/Massacre_des_Italiens_d'Aigues-Mortes

On identifie ici une généralisation abusive. C'est une application inappropriée de la méthode appelée « induction », qui consiste à proposer des lois générales à partir d'informations isolées.

b) De même, on ne peut pas sérieusement conclure à un « remplacement » d'une population par une autre dans un pays au vu de la situation dans quelques quartiers.

Un seul remplacement est incontestable : celui des parents par leurs enfants. Sans lui, l'humanité serait éteinte avant même d'avoir atteint sa deuxième génération ! Ajoutons-y que les enfants ne sont identiques à aucun de leurs parents.

Commentaire sur ces points

Prendre un cas, quelques cas, ou un échantillon et en déduire des caractéristiques pour un ensemble plus large n'est pas forcément illégitime. Mais cela ne peut se faire qu'avec une méthodologie rigoureuse et transparente. C'est par exemple la technique des sondages, qui est une branche des mathématiques, utile quand elle est appliquée sérieusement (voir plus loin)

A3) Et les vaccins ?

Le « problème statistique » des vaccins est très complexe. Il peut sommairement être présenté ainsi :

- on repère des personnes qui, à tort ou à raison, disent en avoir été victimes ; mais peut-on aller de ce particulier au général ?

- les personnes qui ont peut-être été sauvées par leur inoculation ne peuvent pas l'affirmer individuellement tout simplement parce qu'elles l'ignorent ;

- et, même si la vaccination s'avère statistiquement utile, elle pourra être nuisible à certaines personnes. On ne peut déduire d'un bénéfice général qu'il ne serait pas nocif dans des cas particuliers... pour une partie de la population. On ne peut donc aller ainsi du général au particulier.

On se trouve ici devant un aspect de la « malédiction » des statistiques vraies quant aux populations et

fausses quant à l'individu de Léon Schwatzenberg. On ne pourra pas convaincre la famille d'une éventuelle victime d'un vaccin en lui opposant le bénéfice collectif. Surtout si ce vaccin a rapporté énormément d'argent à son diffuseur...

Inversement, le risque éventuel ou potentiel doit-il interdire de mettre en œuvre une politique sanitaire dont les connaissances scientifiques indiquent la nécessité ?

Commentaire sur ce point

La journaliste Lise Barnéoud dans son livre « *Immunisés, un nouveau regard sur les vaccins* » procède à une analyse précise de cette difficulté et propose des solutions. Vaccin par vaccin car on ne peut pas forcément généraliser. Mais aussi en invitant les scientifiques et les politiques à tenir compte, en toute transparence, des alertes signalées et des risques avérés pour certaines populations.
www.radiofrance.fr/franceinter/podcasts/l-edito-carre/vaccination-un-livre-pour-des-choix-eclaires-1236252

En guise de conclusion sur cette partie A

Deux enseignements symétriques :

- une statistique, un ratio, un pourcentage, ne s'appliquent généralement pas de manière identique à tout le monde
- une affirmation juste pour une personne, ou une petite population, n'est pas nécessairement valable pour une population plus vaste ; toute généralisation exige des précautions adaptées à chaque situation. Si la méthode utilisée n'est pas claire, ouverte aux débats, ceux qui s'en méfient auront des raisons, et peut-être raison, d'être sceptiques.

B) Les dangers de la moyenne

« Le statisticien, c'est celui qui a la tête dans le four et les pieds dans le congélateur et qui dit : en moyenne la température est agréable »

Attribué à Edgar Faure [mais il faudrait plutôt dire : le mauvais statisticien, ou la mauvaise statisticienne...]

B1) La réunion de millionnaires... et d'un autre qui influence fortement la moyenne

Une association de millionnaires réunit son conseil d'administration. Cette instance comprend 4 membres.

Trois d'entre eux sont déjà arrivés. Ils sont, respectivement, titulaires des fortunes suivantes ;

- 1.100.000 de dollars pour le « plus pauvre »

- deux millions de dollars pour les deux autres

La « fortune moyenne » de ces trois associés est de $(1,1 + 2 + 2) / 3 = 5,1 / 3 = 1,7$ million de dollars.

Ce qu'on appelle la médiane, c'est la fortune qui se situe au milieu : la moitié des associés a une fortune supérieure ou égale, la moitié a une fortune inférieure ou égale. La médiane est ici de 2 millions de dollars. Notons que la moyenne est ici inférieure à la médiane. Mais cela va changer, et pas qu'un peu, dans la suite de cet exemple !

Car entre dans la pièce le quatrième membre du conseil. Sa fortune est évaluée à 1.000 millions. C'est le milliardaire du groupe

On peut vérifier que la médiane ne change pas : c'est toujours 2 millions de dollars.

Mais que devient la moyenne ? Elle est désormais égale à $(1,1 + 2 + 2 + 1000) / 4 = 251,275$ millions de \$.

Elle est désormais largement supérieure à la médiane. Elle a été multipliée par 148 du fait d'un seul individu.

Commentaire sur ce point

Non seulement la moyenne ne peut à elle seule résumer toute une population mais cet exemple montre qu'elle peut être très éloignée de la personne qui se situe au milieu, l'individu « médian ».

B2) Variante : Le salaire moyen n'est pas celui qu'on pourrait croire

Soit une entreprise de 10 salariés.

9 d'entre eux gagnent 1 500 euros par mois et le patron, lui aussi salarié, 6 500 euros.

Le salaire médian est égal à 1.500 euros.

Calculons la moyenne, c'est le résultat de la division de la masse salariale par le nombre de salariés.

Ici, la masse salariale totale est égale à : $(9 \times 1.500) + 6.500 = 13 500 + 6 500 = 20 000$ euros

Le salaire moyen est donc égal à $20.000 / 10 = 2 000$ euros.

Commentaire sur ce point

Le salaire « médian » correspond sans doute assez bien à ce que l'on entend généralement par « salaire du salarié moyen ». En effet, c'est celui qui partage la population en deux ensemble de même taille : ceux qui gagnent moins et ceux qui gagnent plus. Ici le salaire médian est de 1.500 € alors que le salaire moyen est donc de 2.000 €.

B3) Attention aussi aux évolutions « moyennes » et aux « effets de structure »

a) Reprenons l'exemple précédent et admettons maintenant que le seul salaire du patron augmente de 20%

Soit une augmentation de $6.500 \times 0,2 = +1 300$ euros.

Les autres salaires stagnent. La masse salariale totale de l'entreprise, qui était de 20.000 € est désormais de 21.300€ soit une augmentation de 6,5 % $(100 \times 1.300 / 20.000)$

Le salaire médian, lui, n'a pas bougé. C'est toujours 1.500 €. Dire que « les salaires ont augmenté », du fait que le salaire moyen est désormais plus élevé, est ici radicalement faux

b) Imaginons maintenant une entreprise où les effectifs et les salaires annuels évoluent ainsi

	2023	2024
Nombre de cadres	2	3
Nombre d'autres salariés	10	9
Salaire des cadres	50000	50000
Salaire des autres salariés	20000	20000
Salaire moyen	25000	27500
Évolution 2024/2023		
Du salaire des cadres		0
Du salaire des autres salariés		0
Du salaire moyen dans l'entreprise		10%

La moyenne a crû de : $100 * (27.500-25.000)/25.000 = 10 \%$. Pourtant, le salaire des cadres est resté identique ainsi que le salaire des autres salariés de l'entreprise. Que s'est-il passé ? Un 3ème cadre est arrivé alors qu'un autre salarié est parti. La structure de l'emploi a été modifiée.

Commentaire sur ce point

C'est ce qu'on appelle l'effet de structure. Il est ici visible mais ce n'est pas toujours le cas. C'est sans doute là un des principaux « pièges de la statistique », une source de nombreuses erreurs d'interprétation. Y compris de bonne foi par des spécialistes pourtant compétents. Voir le « bouquet final » page 23.

B4) Quelles pondérations a-t-on employées ?

Reprenons ici cet exemple imaginaire, cité dans le n° 34 de la « Lettre Blanche » de l'association Pénombre, consacrée aux palmarès. À la suite de douze épreuves sportives, les pays A, B et C ont obtenu les médailles d'or, d'argent et de bronze comme l'indique le tableau. Quel pays a, finalement, été « le meilleur » ?

Pays	Or	Argent	Bronze	1 Or seul	2 Toutes médailles	3 Pondération
A	7	1	2	7 (1er)	10 (3ème)	25 (2ème)
B	3	1	10	3 (2ème)	14 (1er)	21 (3ème)
C	2	10	0	2 (3ème)	12 (2ème)	26 (1er)

La première et la deuxième colonne reprennent les présentations habituelles des résultats : le nombre de victoires (médailles d'or), le nombre de podiums (Toutes médailles) Elles ont toutes deux des inconvénients : ne pas tenir compte des « accessits » dans le premier cas. Considérer toutes les médailles comme équivalentes dans le deuxième. D'où la proposition d'accorder une pondération simple aux résultats : 3 points pour une médaille d'or, 2 points pour l'argent, un point pour le bronze. On voit que, selon la manière de compter, chaque pays pourra se situer au premier, au deuxième ou au troisième rang !

Commentaires sur ce point

On donne, ou non, des poids différents aux données de base dont on dispose pour proposer des synthèses et des palmarès qui, souvent, n'ont pas grand sens.

Commentaire général sur la partie B :

Attention aux moyennes, aux « évolutions moyennes » et aux classements ou palmarès, donc. Ils résument ou prétendent résumer une situation générale. Un seul chiffre peut éventuellement donner une idée de l'ensemble dont il est issu mais il ne peut que très rarement le décrire valablement à lui seul.

C) L'ambiguïté sur l'objet dont on parle (domaine concerné, pratiques évoquées...)

*Les statistiques, ça vous fait penser à des choses qu'on n'imaginerait jamais autrement.
Mauvaise pente (2005) Keith Ridgway*

C1) Je suis d'une taille supérieure à la moyenne !

Prenons l'exemple d'un individu du sexe masculin de 50 ans. 1m65 est une taille certainement inférieure à celle de ses homologues.

Cet individu plutôt petit qui prétend être plutôt grand nous a abusés en se comparant à l'ensemble de la population. Y compris les tout-petits qui tirent la moyenne... vers le bas !

Commentaires sur ce point

Si on a fait ce calcul d'une taille moyenne de l'ensemble de la population, on l'a peut-être mesurée correctement mais on a présenté un résultat fallacieux. Il ne suffit pas de mesurer avec précision, encore faut-il être d'accord sur ce de quoi on parle.

C2) « Je représente 30 % des électeurs »

Prenons l'élection européenne de 2019.

Il y avait 47.345.328 inscrits, 23.7306.740 votants dont 1.075.566 bulletins Blancs et nuls soit 22.655.174 exprimés. La liste RN a obtenu 5.286.939 voix, ce qui lui a donné 23 sièges sur les 79 attribués à la France.

Selon l'ensemble auquel on se réfère, ce qu'on appelle le « champ », on pourra dire :

- que le RN n'a obtenu que 11% des suffrages (en se référant aux électeurs inscrits)
- qu'il représente 23 % des suffrages exprimés, car plus de 50% des inscrits se sont abstenus ou ont voté Blanc ou nul
- ou qu'il a obtenu 29 % des sièges, en raison du mode de scrutin, qui n'est pas rigoureusement proportionnel

Les trois pourcentages sont « vrais » mais ne se réfèrent pas au même ensemble, au même champ.

En disant qu'il représente 23% des électeurs, ce qui est généralement la manière dont on présente les résultats électoraux, il considère que les abstentionnistes ou les votes blancs « comptent pour du beurre ». Ou encore : qu'ils sont correctement représentés par ceux qui se sont exprimés. C'est ce que suggère la loi électorale mais on en voit les limites. En effet, plus de la moitié des électeurs ne se sentent pas forcément représentés par ceux qui ont été élus ainsi.

Ajoutons en abordant cet exemple sous un autre angle, que parler en pourcentages (ou en valeur relative), ce n'est pas la même chose que de citer des nombres absolus : ici, revendiquer plus de 5 millions de voix, cela a une autre allure que d'être l'élu de 11 % des inscrits.

Commentaire sur ce point :

Avec quelques chiffres relativement simples, on peut délivrer des messages très différents selon les « conventions » auxquelles on se réfère implicitement. Les commentateurs de résultats électoraux le savent bien mais pas forcément tous les citoyens.

C3) Il y a beaucoup plus de personnes qui meurent dans un lit... qu'en tombant dans un précipice en montagne. (signalé notamment par Étienne Klein)

Cela signifie-t-il qu'il est plus risqué de se coucher que de pratiquer la marche en montagne ou l'alpinisme ? Évidemment non !

Le nombre d'événements (ici : les décès) ne doit pas être considéré dans l'absolu mais par rapport aux pratiques considérées (se coucher vs parcourir des lieux dangereux)

Dans le même ordre d'idées : si le % de décès, par rapport au nombre de patients, est supérieur dans l'hôpital A à celui que l'on observe l'hôpital B, cela ne signifie pas nécessairement qu'on y est plus mal soigné. Peut-être A reçoit-il des malades plus gravement atteints que B. Ou les patients qui sont en sérieux danger y restent-ils plus longtemps. On ne peut, pour proposer une conclusion, faire l'économie d'une analyse approfondie.

Commentaire sur ce point

Ici aussi : de quoi parle-t-on quand on énonce des chiffres ? A quel ensemble, à quelles définitions, se réfère-t-on ? Si on ne le sait pas, si on ne le cherche pas, le chiffre ne signifie rien et peut être utilisé pour nous tromper gravement.

C4) Un enfant qui naît aujourd'hui en France est-il gravement endetté ?

Chaque enfant français porterait 30 000 euros de dette publique à la naissance. C'est du moins le résultat d'un calcul qui a obtenu ainsi un important succès de presse. Ce type d'arguments est en effet repris par ceux qui ne voient dans les dépenses publiques qu'un handicap pour le pays.
atlantico.fr/article/decryptage/30-000-euros-de-dette-publique-a-la-naissance-etc—les-maux-de-la-jeunesse-francaise-en-15-chiffres-

Comment a-t-on obtenu ce résultat effrayant et prétendument édifiant ? En divisant le montant de la dette publique brute (un peu plus de 2.000 milliards d'euros à ce moment-là) par le nombre d'habitants du pays (un peu moins de 70 millions).

En réalité, cet enfant moyen n'existe pas : il naît actuellement chaque année en France environ 700.000 enfants dont les patrimoines familiaux sont très différents. Voici encore une moyenne trompeuse.

Et on peut surtout se poser la question suivante : si cet enfant moyen existait et s'il était tenu de rembourser une part de la dette du pays, serait-il en difficulté ?

Non, car cet enfant serait aussi potentiellement détenteur de sa part du patrimoine national. Quelle serait cette part ? On peut faire le calcul en se référant à la publication annuelle de l'Insee :

<https://www.insee.fr/fr/statistiques/7669703>

Selon cette publication, le patrimoine net positif du pays serait de vingt mille milliards d'euros (près de 7 fois le PIB, si on veut accepter cette comparaison et cet ordre de grandeur).

Cet enfant moyen hypothétique, un résident parmi 68 millions, naîtrait donc avec à peu près 300.000 euros (résultat arrondi de la division de 20.000.000.000.000 par 68.000.000). De quoi rembourser dix fois sa dette estimée ici à 30.000 euros.

Commentaire sur ce point :

Quand on rencontre un chiffre frappant et susceptible d'influencer les citoyens et les citoyennes, de justifier ou de dénigrer des politiques publiques, attention à ce qu'il prétend décrire et à la manière dont il a été construit. Le fait que ce chiffre soit abondamment repris n'est malheureusement pas une garantie de sa pertinence...

Conclusion générale de cette partie

On ne peut être spécialiste de tous les domaines qui donnent lieu à des statistiques. Un peu de bon sens peut souvent permettre de débusquer des erreurs ou des mensonges évidents. Mais pas toujours. C'est le débat public contradictoire, fondé sur des méthodes ouvertes qui peut seul aider les citoyens à repérer méprises et duperies. Ce débat public est indispensable : il doit non seulement concerner l'interprétation des grandeurs publiées, mais également, en amont, les conventions et les méthodes de leur élaboration.

Il peut aussi être utile de s'enquérir des intérêts de ceux qui avancent tel ou tel chiffre et de leurs « biais idéologiques » (y compris de l'auteur de ces lignes qui tente d'être honnête mais est forcément subjectif !).

D) Les traquenards des sondages d'opinion

Ils illustrent les trois thèmes précédents, et y ajoutent un problème « technique » spécifique

Semaine épouvantable : pas un seul sondage d'opinion. Tant pis, nous essaierons de deviner tout seuls nos propres intentions. André Frossard

Les exemples surabondent de sondages trompeurs, plus ou moins volontairement, notamment parmi ceux qui prétendent quantifier « les opinions ». En décortiquant un sondage particulier – qui n'est pourtant pas un des pires – on retrouve plusieurs des problèmes soulevés ci-dessus. Sans prétendre ici résumer la Théorie des sondages, on tentera d'en tirer quelques enseignements généraux. .

D1) Unanimité factice : « Les Parisiens » pensent que...

Titre en une du « Grand Parisien » Hauts-de-Seine du lundi 3 juillet 2023 :

« *Pour les Parisiens, il y a encore trop de voitures dans la capitale* »

Cette affirmation découle d'un sondage Opinionway sur un échantillon de 1.048 personnes « représentatif de la population parisienne âgée de 18 ans et plus ».

On assimile ici les Parisiens, implicitement TOUS les Parisiens, à une opinion peut-être majoritaire. Mais si un seul Parisien ou une Parisienne pensait l'inverse, il serait faux de suggérer que « Les Parisiens pensent que... ». En effet, le respect des minorités devrait interdire de lui attribuer cette opinion.

De plus, les moins de 18 ans résidant à Paris ne sont pas pris en compte dans ce sondage, mais peuvent aussi penser... sur la base d'une expérience, par exemple, de cyclistes dans les rues de la capitale !

Commentaire sur ce point :

En analysant d'autres sondages, on pourrait multiplier les exemples imprécis, voire mensongers, en tout cas trompeurs. Dire, par exemple, que « Les Parisiens partent en août » est une manière fautive de traduire le fait qu'ils seraient plus nombreux à partir ce mois-là plutôt qu'à un autre.

Quand on présente, plus ou moins sournoisement, une opinion comme unanime, c'est probablement faux.

D2) Une réponse « plus fréquente » ne définit pas une majorité

Reprenons le sondage déjà mentionné titré « Les Parisiens » pensent que... accompagné du sous-titre *Pour les Parisiens, il y a encore trop de voitures dans la capitale*

À la question « *De laquelle des affirmations suivantes vous sentez-vous le plus proche concernant les déplacements à Paris* », les % de réponses seraient les suivants :

Il faut continuer à réduire la place de la voiture au profit des autres moyens de transport : 41%

L'équilibre entre la voiture et les autres moyens de transport est bon à Paris : 26%

La place de la voiture a été trop réduite à Paris au profit des autres moyens de transport : 31%

Ne sait pas : 2%

Selon les résultats affichés de ce sondage, seuls 41 % accrédiueraient le titre « il y a encore trop de voitures dans la capitale ». Ce n'est pas une majorité absolue de 50 % ! Ce n'est même pas une majorité relative puisque le nombre de ceux qui n'y souscrivent pas, ou ne savent pas, serait donc de 26 + 31 + 2 = 59%.

Un bon point quand même : le % de « ne sait pas » est mentionné. Ici, il n'est que de 2 % Le nombre d'indécis, d'ignorants avoués, de « sans opinion » ou de « non réponses » est généralement très supérieur. Cela peut d'ailleurs constituer un indice du fait que le questionnement n'est pas pertinent ou mal compris.

Si nous affirmons maintenant que 77 % des sondages « oublient » de signaler le % de « ne sait pas », qu'en pensez-vous ? Nous venons nous-mêmes de vous tendre un petit piège : nous ne mentionnons pas la source de ce chiffre. Il est totalement inventé avec éventuellement l'aplomb du « spécialiste ». Aucune étude, à notre connaissance, ne permet d'avancer un tel chiffre. Mais, en tout cas, les commentaires de sondages, très nombreux, qui ne les « sourcent » pas sont *a priori* suspects.

Commentaire sur ce point :

La notion de majorité, issue d'un sondage, est souvent très douteuse. Cette majorité peut être inexistante ou relative (cas qu'on vient de voir) ; elle peut aussi être déterminée par la manière de poser les questions (voir ci-après) ; elle peut enfin être fluctuante, selon les idées dominantes de l'instant ou de l'époque.

D3) Un sondage n'est qu'un sondage, même s'il est bien fait

Reprenons ce sondage Opinionway.

À la question « *De laquelle des affirmations suivantes vous sentez-vous le plus proche concernant les déplacements à Paris* », on a vu que le % de réponses au premier item serait le suivant :

Il faut continuer à réduire la place de la voiture au profit des autres moyens de transport : 41 %

Si l'enquête a été mal faite, le chiffre qui est mis en avant, soit 41 %, peut être entièrement faux.

Par exemple si les enquêteurs se sont très souvent placés au milieu d'une piste cyclable ou dans une rue partagée entre les automobilistes et les piétons. On est certainement plus proche de 100 % que de 41 % d'anti-voitures dans ces populations-là. Un tel sondage serait alors dit « biaisé », il surestimerait l'hostilité à la voiture.

Inversement, si on a interrogé de très nombreux automobilistes (pas forcément « Parisiens » d'ailleurs) coincés dans un embouteillage. Ces « sondés » énervés sont probablement proches de 0 % à souhaiter des mesures plus sévères contre les voitures. On aura alors biaisé le sondage dans l'autre sens en sous-estimant le nombre de ceux qui voudraient encore réduire la place des voitures.

Pour « faire confiance » à un sondage on peut se demander au moins deux choses :

- si ceux qui l'ont réalisé sont compétents ;
- quel est le commanditaire (la société de sondages peut aisément faire pencher les chiffres dans le sens que celui-ci désire)

Mais admettons désormais que ce sondage-ci, qui se dit représentatif, a été bien fait, dans les règles de l'art.

Dans ce cas ils ne seraient **pas** 41 % mais « environ » 41 %... Environ, car ce nombre est affecté d'une marge d'erreur, ce qu'on appelle souvent une fourchette. Cette marge d'erreur est imputable au fait qu'on a interrogé un échantillon et non l'ensemble de la population. On ne peut pas réellement la calculer quand on ne connaît pas en détails la population étudiée. Mais on considère en général qu'un chiffre de 41 % est connu à 3 points près. Environ !

Corollaire : considérons les intentions de vote en faveur d'une candidature (à supposer que ceci signifie quelque chose, par exemple que cette candidature soit effective !...) Ils passeraient de 49 à 51 % en un mois selon un sondage. Il se peut fort bien que le score soit en réalité passé de 52 à 48 % et donc d'une victoire potentielle à une défaite, quand les chiffres publiés affirmeraient l'inverse... de manière bien imprudente (ou manipulatrice).

Commentaire sur ce point :

Un sondage, même bien fait, ne fournit que des approximations, du phénomène observé. L'ordre de grandeur ainsi fourni peut être intéressant, mais ce n'est qu'une estimation.

D4) « Les Parisiens » pensent que... Mais comment leur a-t-on posé les questions ?

Analysons un peu, maintenant, le libellé des questions de ce sondage.

- 1° Il faut continuer à réduire la place de la voiture au profit des autres moyens de transport 41 %
2° L'équilibre entre la voiture et les autres moyens de transport est bon à Paris 26%
3° La place de la voiture a été trop réduite à Paris au profit des autres moyens de transport 31%
4° Ne sait pas 2 %

Il y a un déséquilibre flagrant entre les trois formulations proposées. En effet, les deux propositions « Il faut continuer à réduire » et « L'équilibre est bon » sont deux formes d'approbation de la politique municipale : celle qui a été menée (qui conduirait donc à un bon équilibre) et celle qui est en cours (continuer à réduire...). Quoi qu'on pense de la politique municipale, il aurait sans doute été plus honnête de proposer deux formulations symétriques et peut-être de poser deux questions au lieu d'une

Par exemple, pour le passé : *La place de la voiture a trop été réduite à Paris* contre *La place de la voiture n'a pas été trop réduite à Paris*

Ou encore, pour l'avenir : *Il faut continuer à réduire la place de la voiture à Paris* vs *Il ne faut pas...*

De plus, on peut supposer que si les questions avaient été posées aux « Grand Parisiens » (les banlieusards) et non seulement aux Parisiens et Parisiennes intra muros (de 18 ans et plus) une majorité – une vraie majorité, même probablement très forte – aurait trouvé la politique de la municipalité trop pénalisante pour les automobilistes.

Commentaire sur ce point :

Il y a de nombreux moyens de « tordre », consciemment ou non, des résultats en formulant les questions, ou en les posant dans un ordre tendancieux, en proposant une position moyenne et sans doute attractive pour tous ceux qui n'ont pas une opinion tranchée, etc.

Plus fondamentalement, certains intellectuels (Pierre Bourdieu, Patrick Champagne) ont développé l'idée que certains sondages « fabriquent l'opinion » sous prétexte de l'observer. En particulier, ils posent souvent aux gens des questions qu'ils ne se posent pas. Ou pas encore : sondages préélectoraux bien avant la date...

En guise de conclusions sur cette partie D

Parmi les nombreux enseignements qu'on peut tirer de ce petit exemple :

- une majorité, même absolue, n'est pas l'unanimité,
- une majorité n'est surtout pas la majorité si elle est relative, c'est-à-dire minoritaire !
- une enquête par sondage n'a de signification éventuelle que pour la population étudiée (ici : un échantillon des Parisiens de 18 ans et plus qu'on a pu contacter) et si elle est bien faite
- les résultats, surtout des enquêtes d'opinion, sont très sensibles au libellé des questions, à leur agencement et même à l'ordre dans lequel elles sont posées, et peut-être surtout au moment où elles le sont
- il est souvent intéressant de repérer le commanditaire des enquêtes et le statut de l'organisme qui les a réalisées : les résultats leur seront volontiers favorables. Une société privée de sondages peut, certes, fournir des résultats qui déplaisent à ses clients. Mais elle risque de perdre des marchés si elle leur fournit des résultats jugés décevants ou non publiables.

La statistique publique court moins ce risque, dans la mesure où les sujets étudiés et les méthodes employées ainsi que les définitions sont en général discutées et diffusées publiquement. Mais à condition de ne pas être dépendante de ceux qui la financent (budgets publics et « clients qui peuvent payer »). C'est en principe le rôle de l'Autorité de la statistique publique, du Conseil national de l'information statistique et du comité du label de la statistique publique que de s'en assurer.

E) Les « simples » erreurs de calcul ou les utilisations erronées d'une méthode, d'une notion mathématique inappropriée

*Le loto, c'est un impôt sur les gens qui ne comprennent pas les statistiques.
Anonyme*

E1) Ne pas confondre points et pourcentages !

Admettons qu'un candidat aux élections ait obtenu 200 voix sur 10000 lors d'un premier scrutin et 300 voix, toujours sur 10000, au suivant.
Il est passé de 2% des voix à 3%.

S'il se vante d'avoir progressé de 50%, ce n'est pas vraiment faux puisqu'il est passé de 200 à 300 voix mais c'est fallacieux.

Si on dit qu'il a progressé d'1% on risque aussi de déformer la réalité en, minimisant, dans ce cas, sa progression.

Commentaire sur ce... point :

S'agissant d'une différence entre deux pourcentages, ce qui s'impose ici, il vaut mieux parler de points (3%-2% = 1 point) plutôt que de %. On devrait dire alors que ce candidat a progressé d'un point, ce qui est plus réaliste, plutôt que de 50% ou d' 1% !.

E2) Ne pas confondre corrélation et causalité

a) Des statistiques des années 50 avait mis en évidence le fait que le nombre de décès était particulièrement élevé lorsque la consommation de charbon pour le chauffage était importante.

Pouvait-il en déduire que « le charbon tue » ? C'était bien le cas chez les mineurs de fond et à long terme mais pourquoi cela serait-il vrai dans la population générale et très rapidement ? Il y a peut-être eu des intoxications au monoxyde de carbone. Mais elle ne peuvent suffire pour expliquer une telle mortalité : elle réside dans le fait que, quand les hivers étaient très rigoureux, on observait à la fois une consommation accrue de charbon et des décès, notamment chez les personnes âgées.

Commentaire sur ce point

On parle de corrélation entre ces deux séries statistiques (consommation de charbon d'une part, décès de personnes âgées d'autre part) mais la causalité principale est à rechercher ailleurs : les rigueurs de l'hiver !

b) Les enfants qui ont des grands pieds ont souvent de meilleures notes en mathématiques. Cette corrélation n'est nullement une causalité : les enfants plus âgés ont de meilleures notes et ont aussi de plus grands pieds mais leurs notes ne s'expliquent pas par la taille de leurs pieds !

Commentaire sur ce point :

Cette notion de corrélation, autrement dit quand deux données ou deux séries semblent statistiquement liées, est tout à fait différente de celle de causalité, c'est-à-dire le lien de cause à effet. Ainsi, tenter de démontrer une théorie en additionnant des statistiques peut être trompeur. L'usage, ici, de graphiques et de cartes peut illustrer une étude rigoureuse et des explications rationnelles mais ne saurait s'y substituer.

Là aussi les exemples fourmillent et la mauvaise interprétation nous menace fréquemment : 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 n'est pas parce que deux événements se produisent l'un après l'autre que le premier est la cause du second. Et ce n'est pas non plus parce que deux événements ont lieu en même temps que l'un des deux explique l'autre. C'est souvent évident dans la vie courante : il existe des coïncidences ou des explications profondes différentes des apparences. C'est vrai aussi pour les statistiques.

E3) Les chiffres ne parlent jamais d'eux-mêmes. Il faut réfléchir pour leur « faire dire » quelque chose

Age du détenteur (années 70)	% de TV couleur	% de Noir et Blanc	Total	
Moins de 30 ans		60	40	100
Plus de 30 ans		40	60	100

Interprétation directe

Vers l'âge de 30 ans, les ménages changent leur TV couleur pour du Noir et Blanc

Bonne interprétation :

Les plus jeunes, qui se sont équipés récemment, sont passés directement à la couleur

Les moins jeunes gardent leur poste en Noir et blanc encore quelque temps

% de Femmes Titulaires du permis

Moins de 40 ans	50
Plus de 40 ans	30

Interprétation directe

Vers l'âge de 40 ans, 40% des femmes se voient retirer leur permis

Bonne interprétation :

Dans les anciennes générations, les femmes étaient plus rares à passer le permis

Commentaire sur ce point

Ces exemples fictifs sont flagrants. D'autres, bien réels et plus subtils, pourraient nous égarer sans qu'on s'en aperçoive. L'erreur des « interprétations directes » est ici de ne pas voir que derrière des photographies instantanées de la population, il y a un film : l'évolution des pratiques des générations successives, qui ont abouti à ces situations.

E4) Les 80/20 ne sont pas une loi statistique

20 % des accidents de la route seraient imputables à l'alcool.

Il y aurait « donc » 80 % de risques de provoquer un accident si on n'a pas bu d'alcool avant de prendre le volant, quatre fois plus que si on en a ingurgité. Ce serait une application de la fameuse loi (??) des 80/20.

Commentaire sur ce point

Au-delà de cet exemple extrême et absurde, les 80/20 sont une manière imagée de présenter certains résultats. Par exemple : indiquer que 20 % des ménages concentreraient 80 % des revenus tel pays. C'est plus parlant que d'évoquer un « indice de Gini » (indicateur d'inégalités souvent utilisé mais très abstrait et réducteur). Mais ces 80/20 ne constituent en rien une loi statistique ou une norme de laquelle il conviendrait de se rapprocher.

Conclusion générale de cette partie

Nombreuses sont les chausse-trapes qui guettent l'utilisateur de statistiques. En être au moins un peu averti limite les risques de se faire rouler par des manipulateurs habiles ou par des incapables péremptoires. Cela peut permettre aussi de repérer les nombres auxquels on peut mieux se fier, qui nous fournissent des « ordres de grandeur ».

Conclusion

« Ce qui est simple est toujours faux. Ce qui ne l'est pas est inutilisable. »

Paul Valéry a raison mais il ne faut pas se laisser trop impressionner par ce cruel dilemme.

Peut-on dire que « les statistiques sont fausses » ? Évidemment, mais c'est un exemple du droit à ne rien dire. On a vu un certain nombre de pièges que les statistiques nous tendent. La liste est forcément incomplète ! Ou plutôt non : on a évoqué des pièges que certains emploient pour nous embobiner, avec des statistiques frelatées ou carrément inventées. Ce ne sont pas les stats qui mentent mais, éventuellement, leurs producteurs ou leurs diffuseurs. Ou leurs dissimulateurs !

Certes, il n'est pas possible aux citoyens de décortiquer tous les chiffres qui leur sont proposés, voire imposés. Même si on est agrégé de mathématiques ou statisticien professionnel, on doit bien se contenter, dans ses décisions, de quelques ordres de grandeur, qui suffisent dans la plupart des cas, même si c'est « simple ». Pour ce qui est compliqué, il est quand même possible de se forger des opinions et de prendre des décisions. Comme consommateur, on ne va peut-être pas s'enquérir du prix des tomates dans le monde entier avant d'en acheter une livre. Idem pour les statistiques.

Outre qu'elle n'est pas toujours réaliste, la recherche de la petite bête peut être suspecte. C'est par exemple le fait des « marchands de doute » dénoncés par Naomi Oreskes. Ces falsificateurs, souvent rémunérés par de très grandes entreprises, ont plusieurs méthodes pour nous enfumer :

- monter en épingle des faits qui semblent remettre en cause des résultats scientifiques. Exemple : ironiser sur le réchauffement climatique quand on entre dans l'hiver ; c'est ici une variante de ce qu'on a appelé « passer abusivement du particulier au général » ;
- discréditer, par exemple, une politique publique en ne signalant que ses risques ou ses prétendus inconvénients ; en ajoutant volontiers « chiffres à l'appui » (voir plus loin la citation sur le lampadaire !)
- discuter à l'infini les hypothèses et les observations sur lesquelles se fondent des thèses pourtant bien assurées ; c'est ainsi que les cigarettiers ont prétendu faire avancer la recherche en trouvant toutes les causes du cancer du poumon, gagnant ainsi des années de profits assassins et en suscitant sans doute des millions de décès prématurés.

Mais, finalement, **les statistiques ne sont pas fondamentalement différentes des mots ou des images, qui peuvent tout aussi bien être trompeuses**. Les principales armes contre les propagateurs de fausses nouvelles, de statistiques frelatées ou d'images pipeautées sont connues. Ce sont : l'éducation, la culture générale et le débat démocratique.

Et, même bien informés, les individus isolés ne peuvent pas éviter tous les traquenards qui leur sont tendus, délibérément ou non : une « autodéfense statistique » efficace requiert des interrogations et actions collectives, associatives et syndicales. Ce manuel pourra un peu y contribuer si des citoyens s'en emparent.

Bibliographie commentée

La Politique des grands nombres : Histoire de la raison statistique, Alain Desrosières,
Paris, La Découverte, 2000, 2^e éd. (1^{re} éd. 1993) (ISBN 978-2-7071-3353-3)

Cet ouvrage magistral, devenu un classique depuis sa première publication en 1993, rassemble plusieurs domaines jamais connectés jusqu'ici de l'histoire des sciences et de la politique : il retrace à la fois l'histoire de l'État, des statistiques, des bureaux de l'administration et de la modélisation de l'économie, domaines dont le rapprochement ne s'est fait que très progressivement. Ainsi, la statistique, qui était au XVIII^e siècle la "science de l'État", ignorait alors les probabilités : elles n'y ont été associées qu'au XIX^e siècle. Au fur et à mesure que la "politique des grands nombres" s'enrichit, elle brasse les jeux de hasard, les risques de la vaccination, les assurances sur la vie, la fixation des tarifs douaniers, la fiabilité des jurys, puis, plus récemment, les effets catastrophiques des cycles économiques et les sondages d'opinion, dont l'auteur propose une analyse fort stimulante. En reconstituant les hésitations, les contingences et les controverses qui définissent la "raison statistique", ce livre ne s'adresse pas seulement aux historiens des sciences, aux économistes ou aux spécialistes de science politique, mais veut ouvrir un débat avec le grand public ausculté par ces appareils statistiques.

Chiffres en folie Association Pénombre

www.fnac.com/a314782/Association-Penombre-Chiffres-en-folie-petit-abecedaire-de-l-usage-des-nombres-dans-le-debat-public-et-les-medias

Face aux dérapages et aux manipulations qu'accompagne cette nouvelle religion du chiffre, les spécialistes de l'Association Pénombre ont décidé de réagir à l'usage illicite des chiffres et à la perversion du débat citoyen qu'il implique. De la multiplication des informations chiffrées aux querelles sur le "vrai chiffre", en passant par les chiffres fantaisistes, les nombres sont devenus omniprésents dans les médias et les discours publics. Car l'information chiffrée apparaît comme une preuve d'objectivité et de sérieux. Elle serait même trop sérieuse pour que l'on ennueie le lecteur ou le téléspectateur avec des considérations techniques sur sa validité... Face aux dérapages et aux manipulations qu'accompagne cette nouvelle religion du chiffre, les spécialistes de l'Association Pénombre ont décidé de réagir à l'usage illicite des chiffres et à la perversion du débat citoyen qu'il implique. Dans cet ouvrage, ils livrent le fruit de cinq années d'observation critique, sous la forme d'un abécédaire plein d'humour. De A comme "Adolescents" à Z comme "Zéro", en passant par C comme "Chômage", I comme "Immigration" ou S comme "Santé", ils proposent un florilège élaboré - nourri de très nombreux exemples - des diverses formes du mauvais usage des chiffres : erreurs involontaires, manipulations délibérées, etc. La lecture du livre est facilitée par un astucieux système de repères permettant de naviguer entre les articles en fonction de diverses grilles d'entrée (par thème, par types d'erreurs ou de manipulations). Et surtout par un ton alerte qui n'évite ni les débats de fond ni... les traits d'esprit, comme le souligne la préface du journaliste Philippe Meyer.

NB : l'association Pénombre a été dissoute en janvier 2024 après plus de 30 ans d'activités et de nombreuses publications. Les archives « riches » peuvent et pourront être consultées en divers lieux dont le site www.penombre.org et celui du Musée d'histoire de la justice criminocorpus.org/fr/

Comprendre les statistiques Michel-Louis Lévy

www.fnac.com/a199377/Michel-Louis-Lévy-Comprendre-les-statistiques

Nul ne saurait convaincre, de nos jours, s'il n'émaillait son discours de quelque pourcentage, indice, sondage... quitte à affirmer que, bien évidemment, il ne croit pas aux statistiques ! Mais il en est de l'information statistique comme du reste : elle ne vaut que si l'on sait s'en servir. Comment et par qui sont fabriqués les chiffres ? Que signifient-ils ? Quels sont les pièges à éviter ? Où trouver les principaux renseignements statistiques sur la population, l'emploi, les prix,...

L'indice des prix à la consommation Florence Jany-Catrice

Collection : Repères Éditeur : La Découverte 2019

L'IPC sert pour l'indexation des salaires, des retraites, mais aussi de divers contrats, par exemple les pensions alimentaires. Il est aussi utilisé par les comptables nationaux pour « déflater » des grandeurs macroéconomiques monétaires et fournir des données en termes « réels ».

Mais comment mesure-t-on cet indice des prix à la consommation ? Quelles ont été les réformes qui se sont succédé pour donner forme à l'IPC du XXI^e siècle ?

Cet ouvrage présente l'IPC en s'appuyant sur l'étude des controverses qui ont jalonné son histoire. Situées dans le contexte socio-économique comme dans celui des idées, ces controverses donnent à voir le caractère éminemment conventionnel et politique de l'IPC et, partant, de nombreux autres indicateurs macroéconomiques, tels que la croissance ou la productivité.

Attention statistiques Comment en déjouer les pièges Joseph Klatzmann

www.decitre.fr/livres/attention-statistiques-comment-en-dejouer-les-pieges-9782707126436.html#resume

Si, dans chaque région française, les agriculteurs consomment plus de pommes de terres par personne que les non-agriculteurs, comment se fait-il que, pour l'ensemble du pays ce soit le contraire ? "Très simplement", répondez-vous... quand vous aurez lu l'explication. De nombreux autres exemples de données piégées ou erronées sont cités dans ce livre : comment peut-on "truquer" l'indice des cours de la Bourse ? Pourquoi y a-t-il moins de décès quand les médecins sont en grève ? Pourquoi les comparaisons

internationales des revenus ne signifient-elles rien ? Comment des instituts de sondage peuvent-ils aboutir à des résultats opposés ? Pourquoi les taux de dévaluation sont-ils à peu près tous faux ? Pourquoi votre espérance de vie est-elle plus élevée que ce que disent les statistiques ? Joseph Klatzmann, ancien administrateur à l'INSEE, ancien professeur d'économie rurale à l'Institut national agronomique Paris-Grignon, directeur d'études à l'École des hautes études en sciences sociales et membre de l'Académie d'agriculture de France, démontre que les statistiques méritent d'être lues avec prudence et nous livre les clés nous permettant d'en déjouer les pièges.

Manuel d'autodéfense intellectuelle, Sophie Mazet

Robert Laffont Juillet 2023

Les hommes-lézards dirigent-ils le monde en secret ?

Le concombre espagnol est-il un serial killer ?

La laïcité est-elle, comme le pot-au-feu, une spécialité française ?

Votre supermarché vous connaît-il mieux que vos parents ?

Face à la complexité du monde, cet ouvrage, drôle, original, accessible et intelligent, propose en neuf chapitres une véritable initiation à l'esprit critique. Du discours des complotistes à celui des publicitaires en passant par ceux des politiciens, des scénaristes de séries télé, des pseudo-scientifiques ou des extrémistes de tout poil, il nous invite à décrypter toutes les formes de rhétorique susceptibles de nous influencer.

Composé de neuf chapitres thématiques, assortis d'une « alerte paranoïa » et d'une « boîte à outils », ce manuel propose d'apprendre à décrypter toutes les formes de discours dont la visée est trop souvent de nous manipuler : qu'il s'agisse du discours politique ou du discours publicitaire, en passant par le discours conspirationniste, de l'infotainment, des séries télé, des pseudo-sciences ou des extrémistes de tout poil.

Un livre essentiel, à l'usage des jeunes générations, mais aussi de tous ceux qui souhaitent combattre efficacement langue de bois, intox, raccourcis, amalgames, fausses rumeurs et psychoses en tout genre.

Le grand truquage Lorraine data

Comment le gouvernement manipule les statistiques Ed La Découverte mai 2009

Maîtriser l'information statistique a toujours constitué un enjeu pour les pouvoirs en place. Mais, depuis quelques années, la manipulation des chiffres s'est amplifiée. L'objectif du président de la République est désormais de contrôler au plus près l'information économique et sociale afin de justifier sa politique.

Multipliant les exemples précis, les auteurs mettent à jour les procédés utilisés par le gouvernement : publications sur des thèmes " sensibles " annulées ou reportées, sélection de chiffres censés flatter l'action présidentielle, modification des indicateurs rendant compte de l'action gouvernementale, dénigrement de la qualité des données de ses propres services lorsqu'elles ne lui sont pas favorables, voire démantèlement plus ou moins discret des organismes chargés de la statistique publique.

La volonté collective des auteurs est de donner à tous les citoyens les clés leur permettant une lecture critique des informations concernant la nature et les résultats de la politique gouvernementale, afin de restaurer les conditions d'un réel débat démocratique.

Statactivism Comment lutter avec des nombres. Isabelle Bruno, Emmanuel Didier, Julien Précieux Ed La Découverte

Les statistiques nous gouvernent. Argument d'autorité au service des managers, elles mettent en nombres le réel et maquillent des choix qui sont, en fait, politiques. Le parti pris de ce livre, qui rassemble les contributions de sociologues, d'artistes et de militants, procède du judo : prolonger le mouvement de l'adversaire afin de détourner sa force et la lui renvoyer en pleine face, faire de la statistique une arme critique. L'histoire de cette forme de contestation dont Luc Boltanski indique qu'elle permet de formuler des " critiques réformistes " passe d'abord par un retour sur la longue controverse sur l'indice des prix en France, présentée par Alain Desrosières.

La deuxième partie du livre s'intéresse à la façon dont on ruse, individuellement et souvent secrètement, avec les règles. L'association Pénombre, composée de statisticiens critiques, y présente une fausse interview du brigadier Yvon Dérouillé, qui explique, face caméra, comment tripataouiller les statistiques de la délinquance.

Chiffre Olivier Martin

Anamosa Paris 2023 96 pages 9 euros

Signalé notamment par Suzy Gaidoz, *Le Monde diplomatique*, Juin 2024 page 25

Les chiffres, omniprésents, posés comme donnés, semblent imposer une vérité laquelle nous devrions nous soumettre. Or il est temps de (re)trouver une capacité de les discuter, d'en déchiffrer les significations. Il est temps de se ressaisir de ces objets sociaux, pour retrouver une capacité d'en débattre, un droit de les critiquer et une liberté de les redéfinir.

Chiffres de contamination, mesure de l'intelligence, nombre de chômeurs, score de popularité, montant de la dette publique, indicateurs de performance... Si les chiffres sont omniprésents dans nos sociétés, que nous disent-ils réellement ? De quoi parlent-ils exactement ? Davantage qu'une vérité sur le monde, ils révèlent nos besoins de nous coordonner, de trouver des manières de faire des choix et de disposer de conventions pour nous entendre. Ils nous parlent d'une multitude de choses qu'ils contribuent en permanence à créer. Fruits de l'activité humaine, ils expriment et matérialisent nos choix, nos valeurs, nos conventions : les chiffres sont des objets sociaux et humains, et non des données naturelles s'imposant à nous.

Tout l'objet de cet ouvrage est ainsi de prendre la mesure des dimensions conventionnelles, sociales et politiques des chiffres, en identifiant les enjeux de pouvoirs auxquels ils sont associés. Ce déchiffrement permet de reprendre conscience des choix qui les fondent, de mieux comprendre leurs portées réelles et qu'ils doivent redevenir les objets politiques qu'ils sont en réalité, objets politiques qui doivent être aussi accessibles au débat démocratique. Il faut donc retrouver une capacité à déchiffrer les chiffres, en ne se laissant pas intimider par l'autorité que leur confère leur apparente naturalité ou les pouvoirs qui les promeuvent.

Quelques sites et articles

Société Française de Statistique

www.sfds.asso.fr

Et en particulier son groupe Statistique et enjeux publics

www.sfds.asso.fr/fr/statistique_et_enjeux_publics/467-groupe_statistique_et_enjeux_publics/

Un service public de proximité démantelé - la statistique agricole départementale

éditions Non lieu

sites.google.com/view/statisticiens-espace-rural/accueil

Analyse économique et compréhension des erreurs de prévision Guillaume Chevillon

Département analyse et prévision de l'OFCE Octobre 2005

ofce.sciencespo.fr

Lutter avec les chiffres

Un guide CGT des comptes de l'entreprise et du droit à l'expertise

www.cgt.fr/lutter-avec-les-chiffres

Le Courrier des statistiques

www.insee.fr/fr/information/3622502

Le *Courrier des statistiques* s'adresse au statisticien, qu'il soit débutant ou expert, mais également au citoyen. L'ambition de la revue est d'aborder l'ensemble des grandes problématiques de la statistique publique.

Économie : derrière les milliards, des choix

Ordres de grandeur et comparaisons

analyses-propositions.cgt.fr/economie-derriere-les-milliards-des-choix

Et, en guise de bouquet final , des « biais de sélection » et des regroupements mystifiants

Quand on a affaire à des populations très différentes, hétérogènes, les conclusions qu'on tire pour l'ensemble qui réunit ces groupes peuvent être paradoxales, tendancieuses et tout simplement erronées. Mais on ne s'en aperçoit pas toujours facilement.

Deux exemples de biais de sélection, concernant les vaccins et... le blindage des avions, sont présentés dans cet article ;

<https://theconversation.com/le-paradoxe-de-simpson-illustre-par-des-donnees-de-vaccination-contre-le-covid-19-170159>

Voir aussi (avant le 10 octobre 2028 !) ce court métrage d'ARTE de 9 minutes :

<https://www.arte.tv/fr/videos/107398-002-A/voyages-au-pays-des-maths/> Le paradoxe de Simpson

En particulier : il apparaîtrait « statistiquement » que plus on fume, plus on court vite.

Cette conclusion, évidemment fautive, semble s'imposer si on considère la population dans son ensemble.

Elle est corrigée si on distingue les hommes et les femmes !

Citations choisies

On n'est pas nécessairement d'accord avec toutes. Surtout les deux dernières !

Les chiffres sont aux analystes ce que les lampadaires sont aux ivrognes : ils fournissent bien plus un appui qu'un éclairage. Nombreuses variantes depuis au moins les années 1960.

Cette formulation est celle de Jean Dion *Le Devoir* - 4 juin 1997

Un mort c'est un malheur ; 100 morts c'est une catastrophe ; 10.000 morts c'est une statistique
Attribuée à Joseph Staline, cette idée serait venue de Kurt Tucholsky (1890-1935)

La baisse du chômage : vraie baisse ou baisse statistique ?

Bonne question, entendue au journal de France culture le jeudi 26 septembre 2013 à 12H00

Les statistiques sont une forme d'accomplissement du désir, tout comme les rêves.

Jean Baudrillard

Je ne crois aux statistiques que lorsque je les ai moi-même falsifiées.

Attribué à Winston Churchill

Les statistiques, c'est comme le bikini : ça donne des idées mais ça ne montre pas l'essentiel !

Nombreuses variantes depuis au moins les années 50. Reprise entre autres par Coluche

J'achète ma viande ici Mes tomates ailleurs Et mes herbes en face Sinon, avec leurs statistiques, ils finiront un jour par connaître ma recette de sauce spaghetti.

Le tour du chat en 365 jours (2006) Philippe Gelluck

Les hommes politiques ne connaissent la misère que par les statistiques. On ne pleure pas devant les chiffres.

Henri Grouès, dit abbé Pierre

La retraite, c'est une transition, ce n'est pas un arrêt. Dès que les gens partent à la retraite, les statistiques le montrent, il y a une courbe de vieillissement et d'accélération du vieillissement absolument incroyable.

Contact, l'encyclopédie de la création (émission de TV canadienne).

Quand un homme est enterré vivant, il n'y a que lui qui en soit informé, ce qui fausse à la fois l'optimisme et les statistiques.

Joël de Rosnay

Il y a trois sortes de mensonges: les mensonges, les sacrés mensonges et les statistiques.

André Ruellan Autobiographie I 246 [attribué aussi, entre autres, à Disraéli]

Les faits sont têtus. Il est plus facile de s'arranger avec les statistiques.

Samuel Langhorne Clemens, dit Mark Twain

Indice unique, indice inique

Titre d'un article du journal *Le Monde* consacré à l'indice des prix à la consommation

Je n'aime pas les statistiques qui sont un moyen moderne et scientifique de dire des mensonges sous des apparences de vérités

Alberto Moravia <

Les statistiques peuvent aider à comprendre le monde. A condition de se souvenir qu'il existe

Inspiré du court métrage d'ARTE « Voyage au pays des maths. Le paradoxe de Simpson » .

Par définition les sondages ne se trompent jamais car ils n'ont pas vocation à prédire.

Laurence Parisot. *Les Échos* 16 avril 2007

On peut discuter de tout sauf des chiffres

slogan utilisé lors de la campagne gouvernementale en faveur de la vaccination contre le Covid

Voilà une affirmation peut-être naïve, en tout cas absurde, contre laquelle on pourra brandir ce manuel !