

Statistiques : législation sources méthodologie

Etienne Guidat

eguidat@unistra.fr

2019-20

échanges
espiritualidad
insertion
perspectives
mutualisation
réussite
ouverture
fondation
CHEMISTRY
equation
biology
 $E = mc^2$
RECHERCHE
SYNERGIES
COMPETENCES
pi
TECHNOLOGY
doctorat
cosmopolite
ENSEIGNEMENT SUPÉRIEUR
biotechnologies
axiome
mécanique
management
capitale
droit
excellence
savoirs
wissenschaft
bibliothèques
médecine
tesis
théologie
gravitation
idéaux
connaissances
musica
langage
INTERNATIONAL
solution
HEURISTIQUE
partenariats
HISTOIRE
physique
mécanique quantique
insertion
PLURIDISCIPLINARITÉ
sciences
gravitation
humain
molécule
ambition
quantique
MASTER
cultures
NETWORK

Cadre légal

- Loi n° 51-711 du 7 juin 1951 - Loi sur l'obligation, la coordination et le secret en matière de statistiques
- Loi n° 78-17 du 6 janvier 1978 - Loi relative à l'informatique, aux fichiers et aux libertés
- Loi n° 2008-776 du 4 août 2008 de modernisation de l'économie (article 144)
- Loi n° 2018-493 du 20 juin 2018 relative à la protection des données personnelles, transcription du règlement n° 2016/679, dit règlement général sur la protection des données (RGPD)



Cadre légal

- Loi n° 51-711 du 7 juin 1951 - Loi sur l'obligation, la coordination et le secret en matière de statistiques
- [Conseil national de l'information statistique](#)



- Ne concerne que la statistique publique
- Rassemble des représentants du monde politique (assemblées), de l'administration (ministères), du monde syndical (employeurs et salariés), du monde associatif (familles, journalistes, etc.), du monde universitaire (enseignants et chercheurs) et des personnalités qualifiées - Président nommé par le par Ministre chargé de l'économie
- 7 commissions thématiques + une pour le recensement
- Statue sur :
 - l'inscription à un programme statistique (coordination - pertinence)
 - la confidentialité
 - la qualité méthodologique
 - l'obligation de réponse
- Conçoit, révisé et tient à jour les principales nomenclatures économiques et sociales

Cadre légal

- Loi n° 78-17 du 6 janvier 1978 - Loi relative à l'informatique, aux fichiers et aux libertés - modifiée en janvier 2006

- [Commission nationale de l'informatique et des libertés](#)



- Concerne les secteurs publics ET privés
- Article 1 : « L'informatique doit être au service de chaque citoyen. Son développement doit s'opérer dans le cadre de la coopération internationale. Elle ne doit porter atteinte ni à l'identité humaine, ni aux droits de l'homme, ni à la vie privée, ni aux libertés individuelles ou publiques. »
- Ne concerne que les traitements automatisés de données à caractère personnel (permettant l'identification de manière directe ou indirecte)

Cadre légal

- Les données :
 - doivent être collectées et traitées de manière loyale, licite et transparente
 - sont collectées pour des finalités déterminées, explicites et légitimes
 - sont adéquates, pertinentes et non excessives au regard de ces finalités
 - sont exactes, complètes et, si nécessaire, mises à jour
 - sont conservées sous une forme permettant l'identification des personnes concernées pendant une durée qui n'excède pas la durée nécessaire aux finalités
 - Sont protégées contre la perte ou l'accès par des personnes non-autorisées
- Sauf obligation légale (CNIS), intérêts vitaux de la personne, exécution d'un contrat, le consentement de la personne sur laquelle des données sont collectées est obligatoire
- Les « enquêtés » doivent être informés de tous ces éléments, ainsi que de l'identité du responsable du traitement, du caractère obligatoire ou facultatif des réponses et des destinataires des informations
- Il est interdit de collecter ou de traiter des données à caractère personnel qui font apparaître, directement ou indirectement, les prétendues origines raciales ou ethniques, les opinions politiques, philosophiques ou religieuses ou l'appartenance syndicale des personnes, ou qui sont relatives à la santé, à la biométrie ou à la vie sexuelle de celles-ci

Cadre légal

- La Commission nationale de l'informatique et des libertés
 - les traitements automatisés de données à caractère personnel font l'objet d'une déclaration auprès de la Commission nationale de l'informatique et des libertés
 - Un délégué à la protection des données est désigné dans chaque organisme gérant des traitements
 - Dans certains cas, une autorisation préalable est nécessaire (sûreté de l'Etat, infractions pénales, génétique, biométrie)
 - Moyens d'action : contrôle sur place, sur convocation, en ligne, y compris sous une identité d'emprunt
 - Sanction : mise en demeure puis, rappel à l'ordre, injonction, retrait de certification, amende de 10 à 20 millions ou 2 à 4% du chiffre d'affaire mondial
 - Au pénal : 5 ans de prison ; 300 000€ d'amende

Cadre légal

Droits des personnes :

Toute personne physique a le droit de s'opposer, pour des motifs légitimes, à ce que des données à caractère personnel la concernant fassent l'objet d'un traitement.

Le responsable de traitement met à la disposition de la personne concernée les informations suivantes :

- L'identité et les coordonnées du responsable de traitement et, le cas échéant, celles de son représentant ;
- Le cas échéant, les coordonnées du délégué à la protection des données ;
- Les finalités poursuivies par le traitement auquel les données sont destinées ;
- Le droit d'introduire une réclamation auprès de la Commission nationale de l'informatique et des libertés et les coordonnées de la commission ;
- L'existence du droit de demander au responsable de traitement l'accès aux données à caractère personnel, leur rectification ou leur effacement, et l'existence du droit de demander une limitation du traitement des données à caractère personnel relatives à une personne concernée.

Un organisme producteur principal : l'INSÉE

Missions :

- Éclairer le débat économique et social
 - Collecter et produire
 - Recensement
 - Enquête emploi
 - Enquête mensuelle de conjoncture auprès des ménages
 - Cadre de vie et sécurité
 - Enquête Statistiques sur les Ressources et Conditions de Vie (SRCV)
 - Répertoire national d'identification des personnes physiques (RNIPP)
 - Fichier électoral
 - Traitement des données d'état civil
 - ...
 - Données de la comptabilité nationale
 - Indice des prix à la consommation
 - Toute une batterie d'enquêtes économiques
 - Traitement des données d'autres administrations (URSSAF, TVA...)
 - Gestion du fichier Sirène

Un organisme producteur principal : l'INSÉE

Missions :

- Éclairer le débat économique et social
 - Analyser
 - Études conjoncturelles
 - Souvent spatialisées
 - Prospective
 - Diffuser
 - Les statistiques produites par l'Insee sont à la disposition de tous. Entreprises, administrations, collectivités, mais aussi chercheurs, enseignants, journalistes ou simples citoyens ont accès à l'ensemble de l'information économique et sociale produite par l'Insee et le système statistique public, dans les limites des règles de la confidentialité
 - Publications - Insée première - TEF - Lettre d'information électronique
 - Données détaillées - y compris jusqu'aux réponses individuelles (dans les limites de la loi)
 - Coordonner
 - Secrétariat du Conseil national de l'information statistique (CNIS)
 - Coordination des services statistiques ministériels
 - Enseigner et développer la recherche
 - Le Groupe des Écoles Nationales d'Économie et Statistique (GENES) - L'ENSAE ParisTech (École Nationale de la Statistique et de l'Administration Économique) - L'ENSAI (École Nationale de la Statistique et de l'Analyse de l'Information)

Des services statistiques ministériels et d'établissements publics

- Particularité française
- Services constitués pour moitié de cadres de l'Insee et pour moitié de personnels des ministères
- Principaux services ministériels :
 - SCEES (agriculture)
 - la DARES (travail)
 - le SES (équipement)
 - le SDES (éducation)
 - la DREES (affaires sociales et santé)
 - le SESSI (industrie)
 - le BSEE (Douanes)
 - ...
- Etablissements publics :
 - INED
 - CREDOC
 - CEREQ
 - Observatoire national de la délinquance
 - ...

Des directions régionales

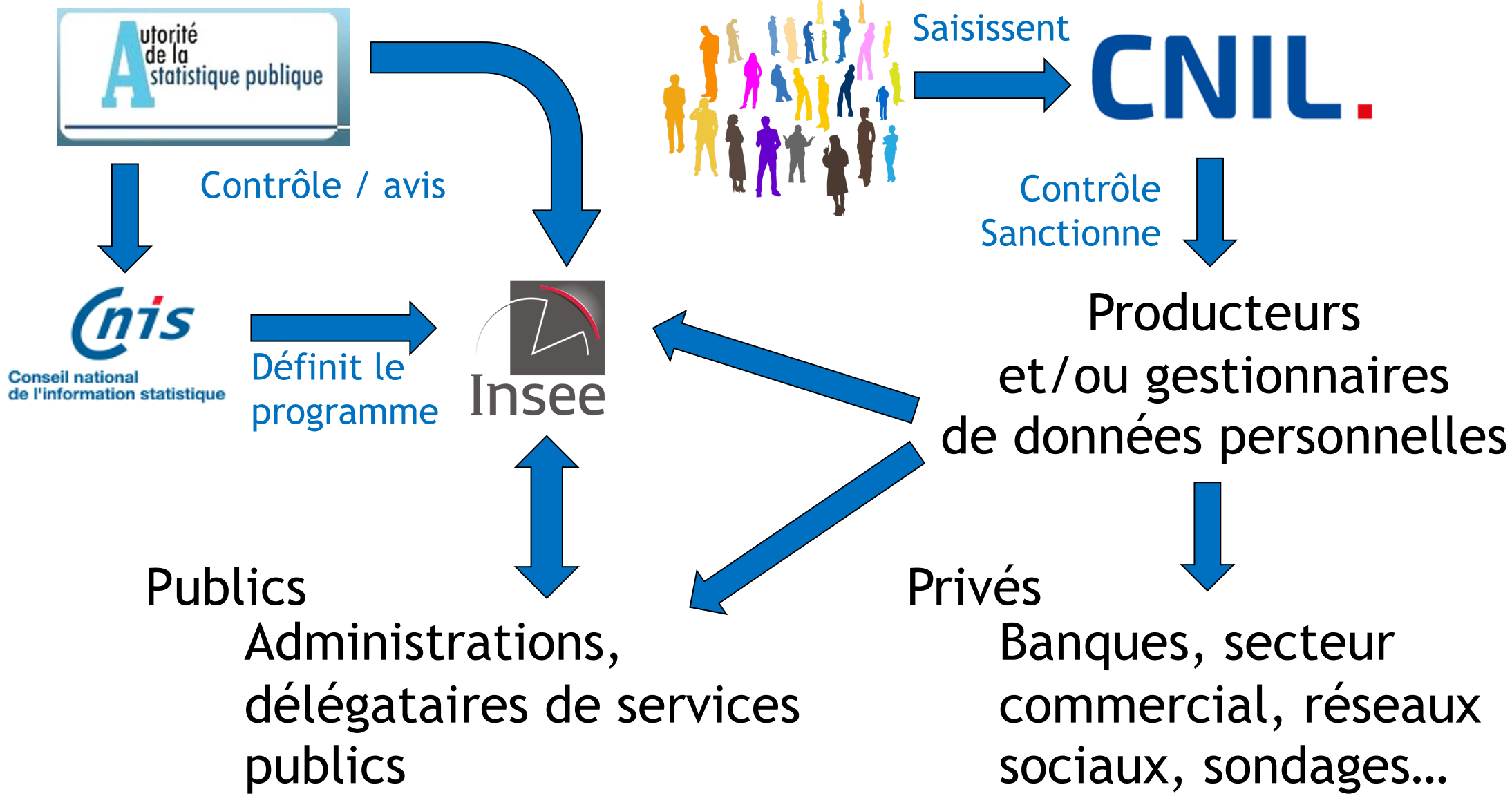
Organisation de la statistique publique en France

Statistiques : sources, législation, méthodologie

En résumé :

Loi de 1951

Loi de 1978 + RGPD



Quelques sites

- <http://www.insee.fr> (sources ; définition ; méthodologie)
- <http://www.reseau-quetelet.cnrs.fr/> (données brutes - accès contrôlé)
- <http://www.data.gouv.fr/> (données brutes)
- <http://www.cnis.fr> (programme statistique ; réglementation)
- <http://www.cnil.fr> (réglementation)
- <https://www.autorite-statistique-publique.fr> (réglementation)
- <http://epp.eurostat.ec.europa.eu>

« On fait dire ce que l'on veut aux chiffres »

Claude Guéant ; *lexpress.fr* ; 08/07/2011

Statistique(s), un terme polysémique

- « Une » statistique = un indicateur, comme la moyenne, le mode, la médiane, l'écart-type... Dans ce cas, la statistique est en fait le résultat d'une méthode statistique
- La collecte des données peut être appelée statistique
- Parfois, les données élémentaires sont appelées statistiques... Exemple, un tableau de données individuelles sera appelé tableau statistique
- La statistique descriptive, ou exploratoire = dans ce cas, la statistique consiste à décrire, en les résumant, et en les simplifiant, des ensembles de données individuelles. Trois principaux outils : graphiques, tableaux, paramètre (ou indicateurs). Historiquement, la première.
- La statistique prévisionnelle = dégager des relations entre différents phénomènes, différentes variables, afin de pouvoir prévoir ou modéliser leurs évolutions
- La statistique inférentielle ou déductive = évaluation d'indicateur sur une population globale, à partir d'un échantillon de celle-ci (sondages)

La statistique descriptive

- La statistique descriptive cherche à décrire quantitativement des ensembles nombreux; elle vise à simplifier (tout en veillant à ce que les simplifications ne soient pas abusives), résumer, synthétiser la connaissance d'un relevé collectif qui échappe, en raison de son étendue et de son instabilité, à l'appréhension directe et individuelle.
- C'est une méthode (plutôt qu'une théorie) ayant pour but de décrire (plutôt que d'expliquer). Si elle est un outil de connaissance, elle facilite l'exercice du jugement, mais elle ne se substitue pas au jugement lui-même (quoiqu'on puisse en dire, les chiffres ne parlent jamais d'eux-mêmes...).

Questions à se poser face à un indicateur

Quelle est la population (ou l'univers statistique)

- C'est l'ensemble des unités ou individus ou éléments sur lequel porte une étude statistique.
- Un individu doit remplir un ou plusieurs critères pour appartenir à une population. Ces critères servent à définir la population. Il est très important de bien définir une population
- Ex : étude de la répartition des cylindrées pour les automobiles en France. La population est le parc automobile français. Il faut bien préciser cette population : voitures immatriculées en France ou roulant sur le sol français... Et à quoi se limite-t-on ? Aux marques françaises ? Aux véhicules de tourisme ? Aux propriétaires de nationalité française ?
- La population française au premier janvier 1996 : il faut indiquer si les étrangers résidant en France sont inclus et comment sont comptabilisés les Français résidants à l'étranger. Il faudra alors préciser la signification de résider.
- Comment définir les personnes employées dans une entreprise au premier octobre 1995 ? Faut-il inclure les travailleurs à domicile, les travailleurs à temps partiel, les travailleurs intérimaires, les stagiaires, les apprentis, les travailleurs « au noir » ? Doit-on comprendre les travailleurs absents pour maladie, congé annuel ou détachement ?...

Questions à se poser face à un indicateur

Suis-je face à une population complète ou un échantillon ?

- Echantillon

- C'est un ensemble d'individus prélevés dans une population déterminée. L'étude statistique ne portera que sur cette fraction d'individus de la population globale. Ca ne dispense pas de bien préciser les conditions d'appartenance à la population globale.
- Raisons pour lesquelles se restreindre à un échantillon ? Coût (185 millions d'euro pour le recensement de mars 1999, soit 3+ euros par habitant), destruction de la population étudiées (crash test).

Quel est le caractère étudié ?

- Caractère

- C'est un trait déterminé, un aspect, une caractéristique, présent chez tous les individus d'une population sur laquelle on effectue une étude statistique : la taille, le poids, la cylindrée d'une automobile.
- La bonne définition du caractère étudié est très importante. Ex : on s'intéresse à la qualification des personnels d'une entreprise. Diplômes ? Diplômes reconnus par l'éducation nationale ? Expériences professionnelles ? Stages ? ...

Questions à se poser face à un indicateur

Quel est le type du caractère étudié ?

- Les caractères qualitatifs
 - Les caractères qualitatifs ou variables nominales ou variables catégorielles ont des attributs dont les différentes modalités ne sont ni mesurables, ni repérables. Elles sont constatées. Avec l'usage de l'informatique, on utilise parfois le terme de données qualitatives. Le caractère se subdivise en catégories ou en modalités de la variable auxquelles seront attachées un effectif et une fréquence. C'est le cas pour le sexe, l'état matrimonial, la qualification professionnelle. Les modalités d'un caractère constituent les rubriques d'une nomenclature. Ce sont des noms ou des étiquettes permettant d'identifier une caractéristique de chaque élément. Il n'est pas toujours possible d'y établir un ordre. Les réponses peuvent être rangées dans une modalité particulière. Un caractère qualitatif peut-être nominal ou ordinal.
 - Les caractères qualitatifs nominaux
 - Une variable nominale décrit un nom ou une catégorie. Les différentes occurrences de la variable sont nominales. Les noms ou les catégories possibles ne suivent pas un ordre naturel. La plupart du temps, la présentation des modalités de la variable ne présuppose aucun ordre, si ce n'est l'ordre alphabétique.

Questions à se poser face à un indicateur

Quel est le type du caractère étudié ?

- Les caractères qualitatifs ordonnés ou variables qualitatives ordonnées
 - Certaines variables appellent naturellement un ordre dans le rangement de leurs catégories (niveau de formation), ... Elles sont repérables selon un type d'échelle plus ou moins légitime. Les catégories pourront alors donner lieu à un codage par les rangs qui ouvrira une autre gamme de traitements possibles proches de ceux des variables quantitatives. Dans le cas d'une nomenclature de formation, le classement est fonction du nombre théorique d'années d'étude nécessaires pour acquérir le niveau de formation. C'est de ce point de vue une variable quantitative repérable, dans quelle mesure le niveau I est-il supérieur au niveau III (comparaison d'un doctorat et d'un BTS). « Excellent » est meilleur que « Très bon », etc.
 - On n'y trouve un certain ordre naturel, mais celui-ci est limité par le fait que nous ne savons pas dans quelle mesure le comportement « Excellent » est meilleur que le comportement « Très bon ».

Questions à se poser face à un indicateur

Quel est le type du caractère étudié ?

- Les variables quantitatives ou numériques
 - Variables numériques discrètes
 - Une variable dont les valeurs sont obtenues par dénombrement est une variable discrète. C'est par exemple le cas du nombre d'enfants. Une variable statistique est discrète ou discontinue lorsqu'elle ne peut prendre que certaines valeurs isolées. C'est le cas du nombre de personnes qui composent un ménage. Un caractère discret peut prendre une infinité de valeurs dénombrables, il peut aussi n'en prendre que quelques-unes : le nombre d'enfants par familles qui est nécessairement un entier.
 - Variable numérique continue
 - Lorsque la variable peut prendre toutes les valeurs à l'intérieur d'un intervalle, la variable est dite quantitative continue.
 - Une variable continue peut être discrétisée
- Modalités
 - Les différentes valeurs ou situations possibles d'un caractère.
 - Au moins au nombre de deux (dichotomique)
 - Exclusif ET exhaustif.
 - Dans le cas d'un caractère qualitatif, les modalités sont les rubriques d'une nomenclature. Dans le cas d'un caractère quantitatif, ce sont les valeurs des mesures de ce caractère. Ces valeurs peuvent être discrètes ou continues

Questions à se poser face à un indicateur

Quel est le type du caractère étudié ?

- Question : de quel type sont les notes que vous obtenez à vos examens ?
- Le type d'un caractère définit les traitements et indicateurs qu'il est possible de lui appliquer pour le résumer
 - Qualitatif : décomptes (fréquence - fréquence relative = effectif des individus d'une population ayant une certaine modalité d'un caractère)
 - Quantitatif : calculs (moyenne, écart-type...)
- Distribution = ensemble ordonné des valeurs prises par le caractère X pour les éléments de l'ensemble étudié

Questions à se poser face à un indicateur

Exercice pratique :

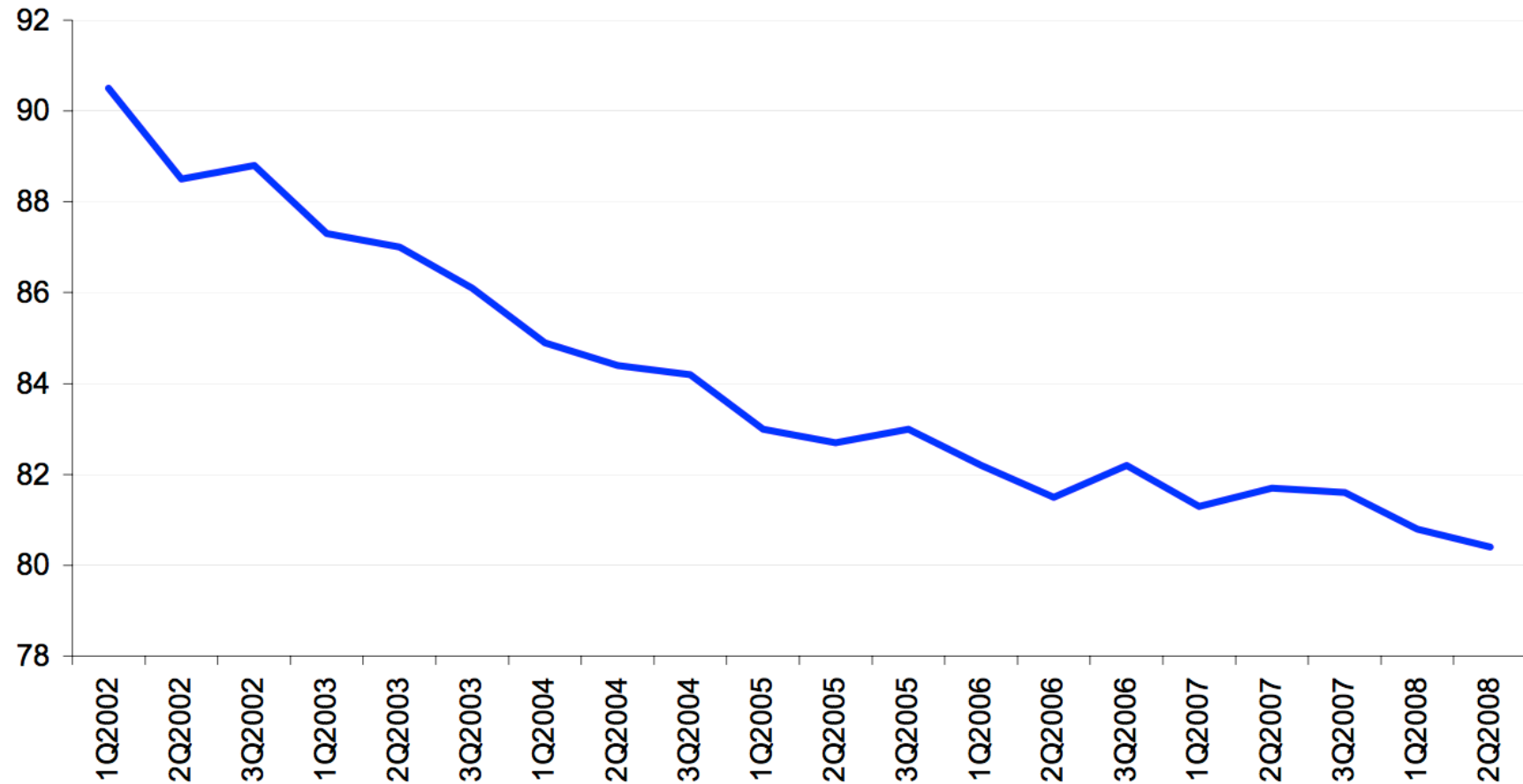
1) « Il y a 25% de chômage chez les 18-25 ans »

Cécile Duflot ; BFM TV ; 14 février 2010

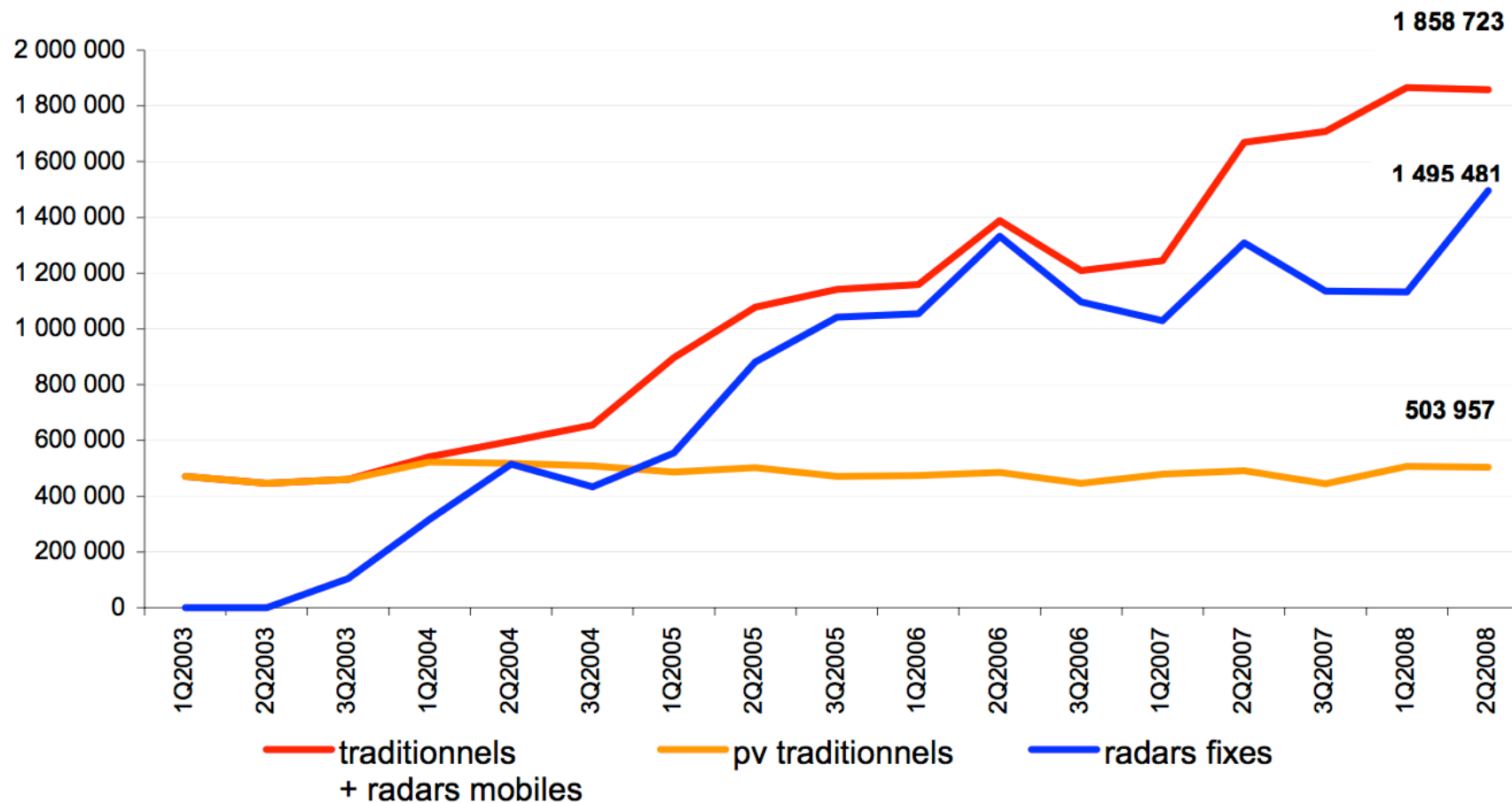
2) Que penser des statistiques produites par Pôle emploi ?

3) L'état 4001 est-il une mesure de la criminalité ?

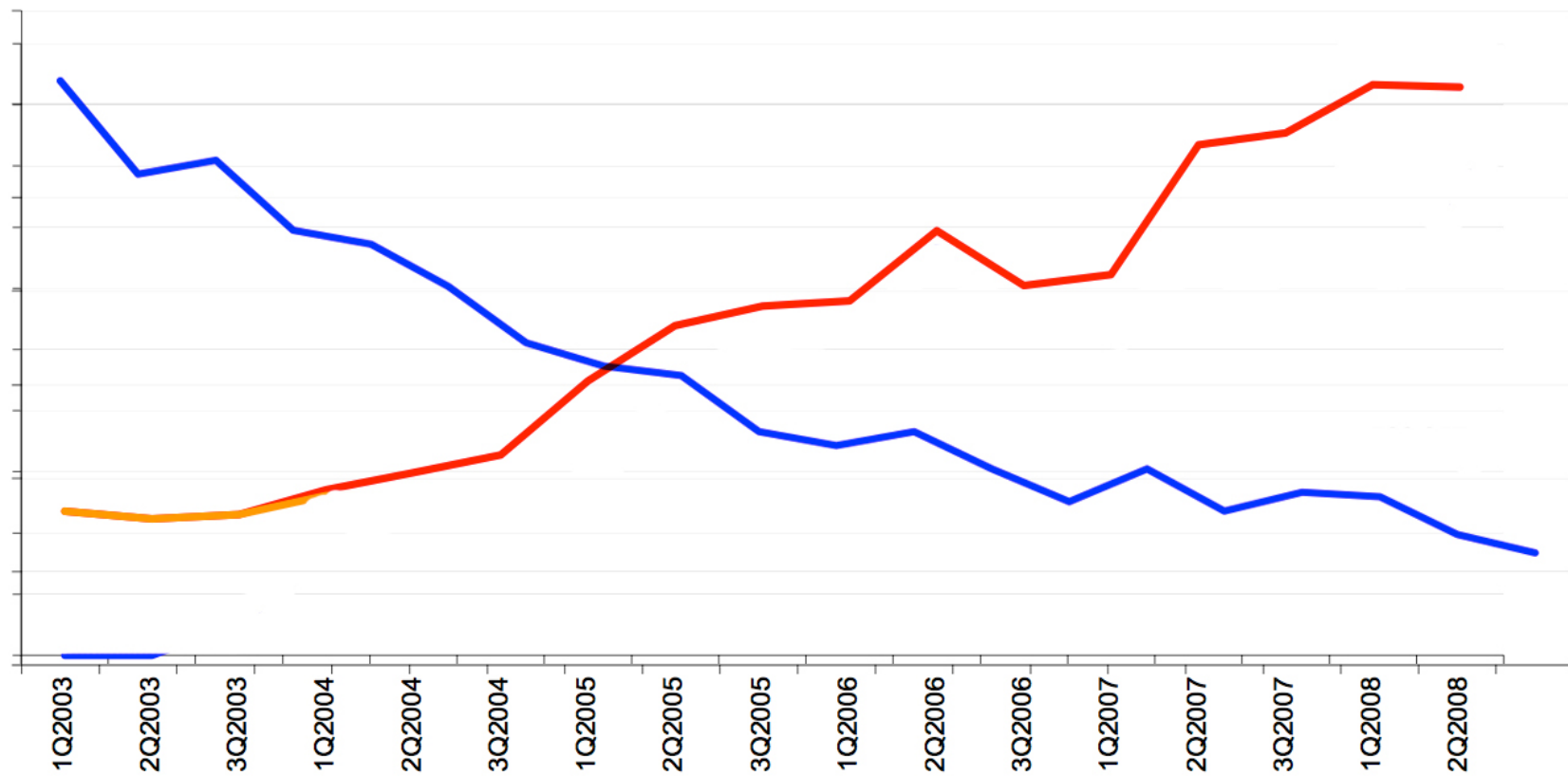
Evolution de la vitesse moyenne des VL



Evolution quadrimestrielle du nombre de PV pour excès de vitesse



Evolution de la vitesse moyenne des VL
Evolution quadrimestrielle du nombre de PV pour excès de vitesse



Paramètres de position d'une distribution

- Mode - Moyenne - médiane
- Objectif : résumer la distribution
- Mode = modalité ayant la plus grande fréquence. Dans le cas d'un caractère quantitatif continu, il faut le discrétiser au préalable, c'est-à-dire établir des classes regroupant plusieurs modalités.
- Médiane = valeur telle que l'effectif de la population dont les modalités sont inférieures à la médiane soit égale à l'effectif de la population dont la modalité sont supérieures à la médiane
- Moyenne (arithmétique) = C'est la somme de toutes les valeurs prise par un caractère, divisé par l'effectif de la population
- Médiane et moyenne ne concernent que les caractères quantitatifs

Paramètres de dispersion d'une distribution

- Etendue - Quartiles - Variance - écart-type
- Objectif : donner une indication sur la similitude ou de la dissimilitude des individus d'une population
- Étendue :
 - Objectif : donner une idée de la « largeur » d'une distribution
 - Différence entre la modalité maximum et la modalité minimum d'un caractère.
- Intervalle interquartile :
 - Q1 et Q3, analogie avec la médiane (Q2)
 - Intervalle interquartile = $Q3 - Q1$
 - Regroupe 50 % de la population
- Variance et écart-type :
 - Objectif : indiquer si les valeurs des modalités des individus sont concentrées autour de la moyenne ou dispersées
 - La somme des écarts à la moyenne est nulle, par définition
 - On calcule donc la moyenne des carrés des écarts à la moyenne = variance. Pas très intuitif => écart-type = racine de la variance => on se retrouve dans l'unité d'origine

Les sondages

- Un sondage, c'est essayer d'obtenir des informations sur une population en observant qu'une partie de ses individus.
- Raisons les plus courantes : coût, rapidité et destruction des individus observés. D'une certaine manière, aussi, meilleures mesures.
- Comme pour une étude exhaustive, il faut définir la population étudiée, l'élément statistique et le ou les caractères étudiés
- Deux grands types de sondages :
 - Aléatoires ou probabilistes
 - Quotas

Les sondages

- Sondage aléatoire : le hasard garantit la qualité de la mesure
- On appelle les sondages probabilistes ceux où tous les éléments statistiques ont la même chance ou, au moins, une chance d'être choisis. Ce sont les seuls qui permettent de mettre en œuvre les méthodes d'analyse statistiques qui permettent d'évaluer la qualité d'un indicateur relever sur un échantillon c'est-à-dire, de maîtriser le risque qu'on a de se tromper
- Exhaustivité de la base de sondage
 - Une première condition pour que tout individu ait une chance d'être choisi est de disposer d'une base de sondage complète. C'est à dire une liste complète de tous les individus de la population que l'on souhaite sonder
 - Dans de nombreux cas, ce n'est pas un très gros problème :
 - tests de qualité industriels
 - cas où la population humaine est maîtrisée et ne peut choisir de ne pas ou peu choisir être sondée (hôpital, entreprise, administration...)
 - En revanche, dans le cas de sondage d'opinion sur la « population française », c'est bien plus délicat d'obtenir une base de sondage complète :
 - Recensement
 - Annuaire : fut mais n'est plus

Les sondages

- La seconde condition est que chaque individu ait avoir la même chance d'appartenir à un échantillon.
- Deux principaux types de non-équiprobabilité :
 - logique d'affinité liée au dispositif de collecte (homme-femme ; personnes peut appréciées = clochards)
 - biais de sélection, liés à la population, aux individus. Certains individus ont intrinsèquement moins de chance d'être tirés. Exemple : Pas là quand l'enquête se passe, refuse de répondre aux enquêtes, adore les enquêtes...
- Si la probabilité d'être tiré au sort n'est pas la même pour tous les individus, on peut encore faire un sondage de type probabiliste si on connaît la probabilité qu'a chaque individu d'être tiré. On doit alors donner aux individus un poids inversement proportionnel à leur probabilité d'être tiré au sort.
- Dans le cas où la base est incomplète, certains individus ont des probabilités nulles d'être choisi.
- Si la probabilité d'être tiré au sort est totalement indépendante du phénomène qu'on cherche à mesurer, tout cela n'est pas important, mais ce n'est jamais le cas.
- Si la base n'est pas exhaustive, ou si on ne connaît pas la probabilité d'un individu d'être choisi, on sort du cadre probabiliste et donc du cadre scientifique. C'est le cas de tous les sondages d'opinion réalisés par les instituts de sondage privés.

Les sondages

- Le paramètre (l'indicateur) devient variable
- Comme on peut obtenir plusieurs échantillons, le paramètre qu'on mesure varie. Les variations de ce paramètre sont ce qu'on appelle les fluctuations d'échantillonnage.
- La moyenne de toutes les valeurs possibles du paramètre donne la valeur exacte du paramètre que l'on cherche à estimer. On dit que son espérance mathématique est la valeur à estimer.
- Rappel sur les dénombrements : combien d'échantillons de 3 individus peut-on faire dans une population de 10 personnes ? 720. C'est à dire, 720 sommes, moyennes, proportions, écarts-types possibles.

- La variance de l'estimateur de la moyenne est égale à : $\left(1 - \frac{n}{N}\right) \times \frac{s^2}{n}$

- Quand N est grand : $\frac{s^2}{n}$

Les sondages

- Cas particulier d'une proportion : 0 si une condition n'est pas respectée ; 1 si elle l'est.

- La proportion est la moyenne de la variable :
$$P = \frac{\sum p_i}{N}$$

• $s^2 = PQ$ (où $Q = 1-P$)

- D'où, la variance de l'estimateur de la proportion :
$$\left(1 - \frac{n}{N}\right) \times \frac{pq}{n}$$

- et, quand N est grand :
$$\frac{pq}{n}$$

- Tout cela pour montrer que : la variation de l'estimateur ne dépend presque que de la taille de l'échantillon et de la variance de la variable étudiée. Mais très peu du taux de sondage.

Les sondages

- Il est possible de démontrer que la distribution de l'estimateur de la moyenne suit une loi normale à partir du moment où l'échantillon est assez grand (en pratique, environs 30 éléments). Avantage, cette loi est très connue, il existe des tables de ses différentes valeurs.
- Cela permet de calculer des intervalles de confiance :

1. Cas de la moyenne

L'intervalle de confiance d'une moyenne est donné par la formule :

$$\bar{x} - u\sqrt{\frac{s^2}{n}} \leq \bar{X} \leq \bar{x} + u\sqrt{\frac{s^2}{n}}$$

Valeurs de u :

Intervalle à 90% u = 1,65

Intervalle à 95% u = 1,96

Intervalle à 99% u = 2,58

2. Cas d'une proportion

$$p - u\sqrt{\frac{pq}{n}} \leq P \leq p + u\sqrt{\frac{pq}{n}}$$

à retenir :

$$p - 2\sqrt{\frac{pq}{n}} \leq P \leq p + 2\sqrt{\frac{pq}{n}}$$

Intervalle de confiance à 95% (en plus ou en moins)

Taille de l'échantillon	Proportion					
	5	10	20	30	40	50
100	4,3	5,9	7,8	9,0	9,6	9,8
200	3,0	4,2	5,5	6,4	6,8	6,9
500	1,9	2,6	3,5	4,0	4,3	4,4
1000	1,4	1,9	2,5	2,8	3,0	3,1
10000	0,4	0,6	0,8	0,9	1,0	1,0
20000	0,3	0,4	0,6	0,6	0,7	0,7

Les sondages

- Sondages par quotas
 - Ils ne respectent aucune règle des sondages probabilistes.
 - Ils ne sont PAS représentatifs. Le hasard n'est absolument pas respecté.
 - Il est impossible d'évaluer leur précision.
 - Les quotas sont un plus mais en aucun cas une garantie.
 - En fait, il y a toujours biais de sélection. Ce n'est pas grave si la(les) variable de sélection sont indépendante de la variable mesurée. C'est rarement le cas.

Discrétiser une variable continue

- Discrétiser = découper une série statistiques continue en classes (toujours l'objectif de simplifier)
- Trouver un compromis entre :
 - résumer au mieux la distribution (plein de classes)
 - Faciliter la compréhension (peu de classes, 5 à 10 en pratique)
- Il y a une part de subjectivité
- Conserver la forme de la distribution
- Conserver la dispersion :
 - Maximiser l'hétérogénéité entre les classes
 - Maximiser l'homogénéité à l'intérieur des classes
- TRES IMPORTANT : nommer les classes

Discrétiser une variable continue

- Quelques méthodes :
 - Classes d'amplitudes égales
 - Classes basées sur la moyenne et l'écart- type
 - Classes d'effectifs égaux

Représentations graphiques

- Simplicité :
 - ne pas faire des graphiques par plaisir
 - l'aspect visuel doit apporter un plus
 - si un graphique devient trop compliqué, c'est qu'il faut faire un tableau
- Attention aux histogrammes et autres graphiques où une valeur est représentée pas une aire !

Méthodologie - Excel

Statistiques : sources,
législation, méthodologie

The screenshot shows the Microsoft Excel interface with a table of keyboard shortcuts. The table is organized into columns A through E. Row 8 is highlighted. An arrow points to cell D4, and another arrow points to cell A16.

	A	B	C	D	E
1	Flèches	Déplacement			
2	Flèches + Maj	Sélection			
3	Ctrl + Souris	Sélection disjointe		Double-clic	
4	= + clic	Entrer une référence		ou cliquer-glisser ici pour ajuster	
5					
6	Cmd + Flèche	Déplacer la sélection jusqu'à la prochaine cellule vide/pleine			
7	Cmd + Flèche + Maj	Etendre la sélection jusqu'à la prochaine cellule vide/pleine			
8	Maj + Espace	Sélectionner la ligne courante			
9	Ctrl + Espace	Sélectionner la colonne courante			
10	Cmd + A	Sélectionner la feuille			
11					
12	Double-clic sur le "carré bleu"	Recopie la cellule vers le bas jusqu'à qu'une cellule vide apparaisse dans la colonne immédiatement à droite			
13	Alt + Cmd	Retour chariot dans une cellule			
14	' en début de cellule	Change un chiffre ou une formule en texte			
15					
16		Double-clic ou cliquer-glisser ici			
17					
18					
19					

Méthodologie - Excel

Statistiques : sources, législation, méthodologie

base-cc-logement-2010-Alsace.xlsx

Rechercher dans la

Accueil Mise en page Tableaux Graphiques SmartArt Formules Données Révision Développeur

Modifier Police Alignement Nombre Format Cellules

Remplir Arial 10 Coller Effacer Renvoyer à la ligne automatiquement Standard Mise en forme conditionnelle Normal_Clas... Normal Bon Insatisfaisant Neutre Avertissement Calcul Cellule liée Entrée Lien hypertexte Remarque Sortie Insérer Supprimer Format

	A	B	C	D	E	F
1	=A6	Cellule				
2	=B8:C11	Zone				
3	=B8;C11	Cellules disjointes				
4	=B8;C11;C13:C18	Zones disjointes				
5						
6	Texte	Constante alphanumérique (gauche)				Constante
7		16 Constante numérique (droite)				
8	=C3+D3	Formule (affiche 12) - Barre de formule	5	7	9	Dynamique
9	=SOMME(C3:E3)	Fonction (affiche 21) - Double-clic				
10	=SOMME(Mazone)	Fonction (affiche 21) - Insertion -> Nom -> Définir / Zone de nom				
11						
12	=A6	Référence relative = "mise à jour" lors d'un copier/coller				
13	=\$A6	Référence semi-relative : colonne fixe				
14	=A\$6	Référence semi-relative : ligne fixe				
15	=\$A\$6	Référence absolue : fixée lors d'un copier coller				
16						

Références Clavier Fonctions Filtres et tris Sitadel2012 Sitadel2011 COM-logement Liste des variables Définitions Formules Conseils utilisation discrétisation Graphiques Analyses multivariées Exercices 1 +

Mode Normal Prêt Somme=0

Méthodologie - Excel

Statistiques : sources, législation, méthodologie

base-cc-logement-2010-Alsace.xlsx

Arial 10

Accueil Mise en page Tableaux Graphiques SmartArt Formules Données Révision Développeur

Modifier Police Alignement Nombre Format

Remplir Arial 10 A A- Renvoyer à la ligne automatiquement Standard % 000 Mise en forme conditionnelle

Coller Effacer G I S Fusionner

A15 fx NB

	A	B	C	D	E	F	G	H	I
1									
2	Syntaxe d'une fonction :	signe "=" + "motclef" + "(0 ou plus arguments séparés par des ;)"							
3		=somme(d1:d2)							
4		Insertion -> Fonctions...							
5									
6			Exemples						
7	GAUCHE	Extraire n caractères à gauche d'une chaîne	Gau	Gauche					
8	DROITE	Extraire n caractères à droite d'une chaîne	ite	Droite					
9	STXT	Extraire n caractères dans d'une chaîne	aîn	Chaîne					
10	NBCAR	Concaténation de chaînes	aîn-ite-Gau						
11	&								
12									
13	Décomptes								
14	NBVAL	Donne le nombre de valeurs dans une plage	4	12	A		6	5	
15	NB	Donne le nombre de valeurs numériques dans une plage	3						
16	NB.SI	Donne le nombre de valeurs respectant une condition dans une plage	1						
17			2						
18			2						
19			5						
20									
21	Statistiques								
22	MAX	Donne le maximum d'une plage	5						
23	MIN	Donne le minimum d'une plage	1						
24	QUARTILE	Donne un quartile	2						
25	CENTILE	Donne un centile	4,5						
26	MOYENNE	Calcule la moyenne	2,83333333						
27	MEDIANE	Donne la médiane	2,5						
28	VAR.P.N	Calcule la variance	1,80555556						
29	ECARTYPE.PEARSON	Calcule l'écart-type	1,34370962	1,34					
30	KURTOSIS	Calcule coefficient d'aplatissement	-0,8591716						
31	COEFFICIENT.ASYMÉTRIE	Calcule le coefficient d'asymétrie	0,41807152						

Références Clavier Fonctions Filtres et tris Sitadel2012 Sitadel2011 COM-logement Liste des variables Définitions Formules Conseils utilisation discrétisation Graphiques Analyses multivariées Exercices 1 +

Méthodologie - Excel

Statistiques : sources, législation, méthodologie

The screenshot shows the Microsoft Excel application window. The 'Données' (Data) menu is open, displaying options such as 'Trier...' (Sort...), 'Filtre' (Filter), 'Effacer les filtres' (Clear filters), 'Filtre avancé...' (Advanced filter), 'Grille...' (Table), 'Sous-totaux...' (Subtotals), 'Validation...', 'Table de données...' (Table of data), 'Convertir...' (Convert), 'Consolider...' (Consolidate), 'Grouper et créer un plan' (Group and create a plan), 'Tableau croisé dynamique...' (PivotTable), 'Outils de tableau' (Table tools), 'Données externes' (External data), and 'Actualiser les données' (Refresh data). The background shows a table with columns A, B, and C, and rows 1 through 5. The data in the table is as follows:

	A	B	C
1	LIBGEO	P10_LOG	
2	Achenheim	926	
3	Adamswiller	193	
4	Albé	363	
5	Allenwiller	205	187

The screenshot shows the Microsoft Excel application window with the 'Filtre' (Filter) dialog box open. The dialog box is titled 'LIBGEO' and contains the following options:

- Trier** (Sort):
 - Direction: Croissant (Ascending) / Décroissant (Descending)
 - Par couleur: Aucune
- Filtre** (Filter):
 - Par couleur: Aucune
 - Choisir: (dropdown menu)
 - Rechercher: (search bar)
 - Liste de filtres: (list of filters with checkboxes)
 - Effacer le filtre: (button)

The list of filters includes:

- (Sélectionner tout)
- Achenheim
- Adamswiller
- Albé
- Algolsheim
- Allenwiller
- Alteckendorf
- Altenach
- Altenheim
- Altkirch
- Altorf

Méthodologie - Excel

Statistiques : sources, législation, méthodologie

Excel Fichier Édition Affichage Insertion Format Outils **Données** Fenêtre Aide

Accueil Mise en page Tableaux Graphiques SmartArt Formules

Modifier Remplir Police Coller Effacer

A1 LIBGEO

	A	B
1	LIBGEO	P10_LOG
2	Achenheim	9
3	Adamswiller	1
4	Albé	3
5	Allenwiller	205

Données

- Trier... ⌘R
- Filtre ⌘F
- Effacer les filtres
- Filtre avancé...
- Grille...
- Sous-totaux...
- Validation...
- Table de données...
- Convertir...
- Consolider...
- Grouper et créer un plan
- Tableau croisé dynamique...
- Outils de tableau
- Données externes
- Actualiser les données

A B C D

LIBGEO P10_LOG P10_RP P10_RSEC

Achenheim 926 875

Adamswiller

Albé

Allenwiller

Alteckendorf

Altenheim

Altorf

Altwiller

Trier

Ajouter des niveaux pour trier par : ☒ Ma liste contient des en-têtes

	Colonne	Trier sur	Ordre	Couleur/Icône
Trier par	P10_LOG	Valeurs	Du plus petit au plus grand	
Puis par	P10_RP	Valeurs	Du plus petit au plus grand	

+ - Copier

[En savoir plus sur le tri](#) Options... Annuler OK