QUE SAIS-JE ?

Pourcentages et tableaux statistiques

MICHEL NOVI

Maître de conférences de sociologie
à l'Université de Nice - Sophia-Antipolis

ISBN 2 13 048880 3

Dépôt légal — 1re édition : 1998, août

INTRODUCTION

Les sciences humaines et sociales – la sociologie et la psychologie en particulier –, la démographie et l'économie, les disciplines biologiques ou médicales font un usage intensif des pourcentages. Pourtant, il n'existe pas de Traité des pourcentages ou des proportions ! Sans doute les pourcentages sont-ils la chose statistique la mieux partagée. Calculer un pourcentage, ce n'est après tout qu'effectuer une règle de trois : si x est à 100 comme a est à b alors $x = 100 \times a/b$, valeur qui exprime a en pourcentage de b. Or les règles de calcul concernant la proportionnalité sont une affaire entendue, à tout le moins depuis Euclide et ses « Éléments ». L'origine d'une notion comme le taux se perd d'ailleurs dans la nuit des temps ainsi que l'atteste le Code d'Hamourabi (XVIIIe siècle avant J.-C.) fixant à un cinquième le taux de l'usure.

Ces notions sont si familières qu'elles se prêtent aux traitements métaphoriques les plus divers. Pour un jeu national « 100 % des gagnants ont tenté leur chance », pour les soldes d'un grand magasin « aucun être humain ne résiste à – 40 % »... L'humour de tel messages publicitaires désigne aussi, en contrepoint, bien des abus commis au nom de la statistique en général et des pourcentages en particulier.

C'est qu'il y a, entre le calcul scientifique et la métaphore, la place pour les usages sociaux les plus divers comme l'exposition par les médias de maints résultats d'enquêtes dont l'intérêt ou l'exploitation ne sont pas sans poser problème mais dont la légitimité est imposée par l'évidence du chiffre. La récurrence de la critique est bien attestée par cette citation :

Depuis que la multitude des journaux et des publications de toute espèce répand, chaque jour, un flot ininterrompu d'informations et de documents, la vulgarisation de la statistique s'est étendue sans beaucoup de profit pour elle (André Liesse, *La statistique*, Paris, Alcan, 1904).

Le but de ce livre est de faire le point sur des connaissances et des pratiques souvent jugées trop élémentaires pour être enseignées, mais dont la maîtrise garantit aussi bien la rigueur de l'analyse scientifique que la réception critique des messages sociaux. Son but est également d'aller aussi loin que possible pour le lecteur non spécialiste, de la règle de trois à des méthodes statistiques professionnelles et élaborées.

Le point de vue adopté sera toutefois restreint à celui de la statistique descriptive, par opposition à la statistique inductive ou inférentielle. Ce point de vue correspond, en effet, au premier temps d'une démarche qui a fait ses preuves : observer d'abord pour tenter de généraliser ensuite. Pour reprendre les définitions proposées par Henry Rouanet[1], nous appellerons « descriptive » toute méthode statistique *indépendante* de la taille des groupes étudiés. Ainsi les distinctions – qui caractérisent la démarche inductive – entre *échantillon* et *population* ou entre *teneur* et *portée* des résultats, ne seront pas, sauf exception, appliquées ici. En d'autres termes, on ne se demandera pas si un pourcentage est précis ou si une différence est significative, mais quels sont les moyens, plus ou moins raisonnables, de calculer ce pourcentage ou de mesurer cette différence.

Je tiens à remercier, pour leurs conseils précieux, Sébastien Christian, Valérie Erlich, Alain Frickey, Nausicaa Marlin, Patricia Novi et Jean-Luc Primon.

1. H. Rouanet, B. Le Roux, M.-C. Bert, *Statistique en sciences humaines : procédures naturelles*, Paris, Dunod, 1987.

Chapitre I

POURCENTAGES

Avant d'étudier dans le détail les applications du calcul des pourcentages, nous nous demanderons tout d'abord à quoi tient, techniquement parlant, l'universalité de leur emploi en statistique. Nous préciserons également les définitions et la rationalité de la notion de pourcentage ainsi que des notions qui lui sont apparentées.

I. — Extensivité du calcul des pourcentages

Quelles sont les conditions requises, dans le recueil de l'information, pour qu'un pourcentage puisse être calculé ? Nous allons voir que l'aspect universel du calcul des pourcentages, en statistique, tient à ce qu'il est toujours possible de ramener la mesure des choses à ce que l'on appelle une « échelle nominale ».

1. **Les niveaux de la mesure.** — Mesurer les choses et les phénomènes évoque en général l'idée de nombre et de mesure quantitative. Cependant, à partir de la seconde moitié du XXe siècle, le développement des sciences de l'homme a conduit à distinguer divers types d'échelles de mesure. En simplifiant, on distingue les échelles nominales, ordinales, et numériques.

A) *Les échelles nominales.* — Dans ces échelles, les catégories attribuées aux objets ne sont que des noms, des étiquettes dont la fonction est de classer. Le genre, par exemple, est une échelle nominale composée de deux catégories : masculin, féminin. La nationalité, l'état matrimonial, l'état médical, le secteur d'activité, la religion, le type d'habitat, etc., sont des exemples d'échelles nominales. Ces catégories déterminent donc des classes, par exemple la classe des hommes et la classe des femmes. La seule opération possible est celle qui consiste à dénombrer les objets qui ont été classés dans chacune des catégories, c'est-à-dire à calculer des effectifs par classe.

On dira encore que le genre, par exemple, est une *variable* nominale (ou qualitative). En toute rigueur, une *variable* est une application d'un domaine de définition (individus, par exemple) vers un domaine de variation (échelle).

Sur la figure 1, l'échelle nominale est le domaine de variation (les catégories H et F). La variable – le genre – est la procédure (symbolisée par les flèches) par laquelle on affecte à chacun des cinq individus une catégorie de genre.

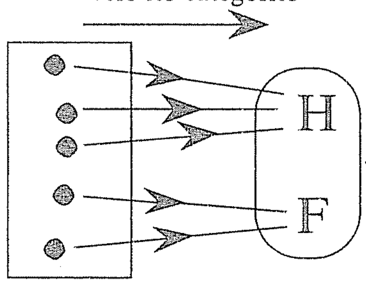

Fig. 1. — Le genre comme échelle de mesure.

La figure 2 montre comment la correspondance, lue dans l'autre sens, induit en retour sur l'ensemble des individus une classification ou « partition ». On peut alors s'intéresser à la (ou les) catégorie(s) la (les) plus nombreuse(s) ; on peut aussi exprimer les effectifs en proportion de l'effectif total. Bref, les échelles nominales constituent le terrain de prédilection du calcul des effectifs et des pourcentages.

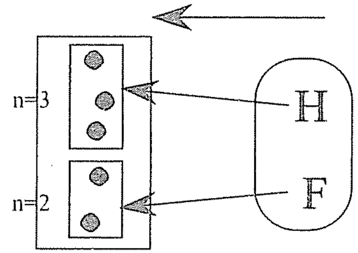

Fig. 2. — La détermination des effectifs.

Ce mécanisme – l'effectif d'une classe est le nombre d'éléments contenus dans l'image inverse d'une catégorie – est à la base du dénombrement et, par suite, du calcul d'indices statistiques.

Cette extension moderne de la notion de mesure peut surprendre : les « échelons » de notre échelle ne sont même pas ordonnés ! Mais il reste qu'avec une échelle nominale on peut toujours dire si deux individus sont équivalents – « égaux » – ou non. Ils sont égaux s'ils sont dans la même classe (ils ont le même genre). Ils sont inégaux s'ils sont dans une classe distincte (leurs genres sont différents).

B) *Les échelles ordinales.* — Dans une échelle ordinale, les catégories assignées par la mesure aux objets sont liées par une relation d'ordre.

Ce type d'échelle est très utilisé dans les enquêtes d'opinion ou les recherches sur les attitudes. On propose en général un objet (une opinion, un homme politique, un énoncé, une situation, un produit...) et on demande à l'enquêté une réponse à choisir sur une échelle, en terme de confiance, d'approbation, d'évaluation, de satisfaction : Êtes-vous : très satisfait ? plutôt satisfait ? etc.)... Il arrive fréquemment que, par commodité, on fasse s'exprimer des sujets sur une échelle de nombres alors que ces nombres ne seront évalués qu'à titre de rangs : ainsi, la douleur physique pourra dans certains cas être mieux exprimée par le sujet sur une échelle arbitraire, de 0 à 10 par exemple, que par une verbalisation.

C) *Les échelles numériques.* — Les catégories attribuées aux objets sont maintenant des nombres, utilisables en tant que tels. En dehors des mesures physiques (masse, temps, longueur, etc.), le revenu, le score à un test, le nombre d'enfants à charge, l'âge, le salaire, la taille d'une entreprise, le chiffre d'affaires, les nombreux indices de l'économie et de la démographie, etc., sont autant d'échelles numériques sur lesquelles certaines opérations mathématiques seront possibles, plus complexes que le dénombrement ou l'ordination.

2. **Passage entre niveaux.** — Il est clair que les divers types de mesures sont hiérarchisés. On parle de « niveaux de la mesure », le niveau le plus élevé étant celui des échelles numériques. Mais rien n'interdit de ramener ces dernières échelles à des échelles ordinales ou nominales. Considérons par exemple les notes données aux candidats d'un examen. On pourra déterminer : au niveau numérique, la moyenne des notes ; au niveau ordinal, la médiane (note dépassée par la moitié des candidats) ; au niveau nominal, le mode (la note obtenue le plus de fois). On pourra dénombrer les étudiants ayant obtenu une note comprise entre deux valeurs x et y, les étudiants ayant moins que x ou plus que y, etc. Comme on le voit, de telles opérations de réduction d'une échelle numérique à des échelles de niveau inférieur ne sont pas nécessairement des dévaluations d'échelle. Elles sont possibles dès que l'on considère que les échelons d'une échelle numérique ou ordinale sont *a fortiori* des catégories, quitte à définir des regroupements en classes. Ainsi un décompte d'effectif sera toujours possible, par intervalles de notes, tranches de salaires, d'âges, etc.

II. — Pourquoi des pourcentages ?

Pour la « pensée statistique »[1], la raison d'être des calculs de proportions tient, d'une part, à la variabilité des phénomènes et, d'autre part, à ce qui fait l'essentiel de la méthode scientifique, la comparaison.

1. D. Schwartz, *Le jeu de la science et du hasard. La statistique et le vivant*, Paris, Flammarion, 1994.

1. **La variabilité des phénomènes.** — La répartition par baccalauréat de 204 étudiants de 1re année de sociologie (Université de Nice, 1995) est la suivante :

Tableau 1. — **En première année de sociologie**

Baccalauréat	*Effectif*
A ou L (littéraire)	46
B ou ES (économique et social)	52
C, D, E ou S (scientifique)	8
Technologique	75
Équivalence	23
Total	204

Ces divers effectifs montrent que les étudiants ne sont pas issus de la même série du bac ! C'est pour cette raison que la série du bac constitue une variable. Il est vrai que l'expression de pourcentages n'ajouterait rien à cette constatation. Le calcul $46/204 \times 100 = 22{,}5$ n'aurait d'utilité que d'éviter la donnée de *deux* nombres, 46 et 204. Dans 22,5 (pour cent), le nombre 100 est, en effet, sous-entendu.

En revanche, un calcul de proportion est nécessaire pour chiffrer la variabilité de l'effectif. Il est tout à fait possible, en effet, de chiffrer la *dispersion* d'une variable nominale. Il suffit de calculer le « rapport de variation »[1] :

$$RV = 1 - n_{\text{mode}}/N$$

où N est l'effectif total et n_{mode} l'effectif de la classe la plus nombreuse (appelée mode ou classe modale).

Dans l'exemple : $RV = 1 - 75/204 = 0{,}63$.

Cet indice de dispersion, on ne peut plus simple, mesure la dispersion de l'effectif hors de la classe modale. Pour K modalités, RV varie entre $1 - N/N = 0$ (dispersion minimum : l'effectif est concentré dans une seule classe) et $1 - (N/K)/N = (K - 1)/K$ (dispersion maximum des N unités dans les K classes).

1. H. M. Blalock Jr., *Social Statistics*, Paris, McGraw-Hill, 1979.

RV ne tient compte que du seul effectif de la classe modale. Deux autres indices de dispersion sont utilisables : l'indice de diversité et l'entropie.

A) *L'indice de diversité.* — La dispersion de la variable « série du bac » peut être évaluée par :

$$D = 1 - \Sigma\,(n_i/N)^2 \quad i = 1\ \dots\ K$$

n_i désignant les effectifs obtenus pour chaque modalité et N l'effectif total. Dans l'exemple, on a :

$$D = 1 - (46/204)^2 - (52/204)^2 - \dots - (23/204)^2 = 0{,}735$$

indice qui varie, comme RV, entre 0 et Dmax = (K – 1)/K et qui peut être interprété comme la probabilité pour que deux étudiants tirés au sort ne relèvent pas de la même série.

Imaginons, en effet, une urne où chacun de nos étudiants ait sa série inscrite sur une boule. Tirons au hasard une boule parmi les 204 et relevons la série. La probabilité que la série observée soit littéraire est 46/204 = 0,225. Remettons la boule dans l'urne et répétons l'opération. La probabilité que les deux séries observées soient littéraires est :

$$(46/204)\,(46/204) = 0{,}225 \times 0{,}225 = 0{,}051.$$

De la même manière, la probabilité que les deux séries soient B ou ES est 0,065. Au total, la probabilité que les deux séries soient identiques est :

$$0{,}051 + 0{,}065 + 0{,}001 + 0{,}135 + 0{,}013 = 0{,}265.$$

A l'opposé, la probabilité pour que les deux séries soient différentes est 1 – 0,265 = 0,735, indice qui mesure la dispersion dans les séries. La valeur maximum étant (5 – 1)/5 = 0,8, on peut « normer » le coefficient en calculant $D_{nor} = D/D_{max} = 0{,}735/0{,}8 = 0{,}919$. D_{nor} varie maintenant entre 0 et 1 et ne dépend plus du nombre K de modalités.

B) *L'entropie.* — La théorie de l'information fournit, par ailleurs, une autre excellente mesure de la dispersion (ou de l'incertitude) avec le calcul de l'entropie :

$$E = -\,\Sigma\, p_i \log_2(p_i) \quad \text{où } p_i = n_i/N \quad i = 1\ \dots\ K$$

E varie entre 0 et Emax = $\log_2$(K). On a, pour la série du bac : E = 2,056, $E_{max} = \log_2(5) = 2{,}322$ et pour le coefficient normé : $E_{nor} = E/E_{max} = 0{,}886$.

D et E ont des comportements tout à fait similaires et nous n'utiliserons que le premier indice par la suite. Pour la notion d'information et ses rapports avec l'analyse statistique, on pourra consulter Jean-Marie Faverge, Jean-Paul Benzécri, Michel Volle[1].

2. **Rapporter pour comparer.** — C'est en proportions qu'il convient en général de mener les comparaisons. Le tableau 2 nous donne, en milliers, les effectifs des cadres d'entreprise et des techniciens lors des deux derniers recensements :

Tableau 2. — **Entre deux recensements**
Effectifs exprimés en milliers.

	1982	*1990*
Cadres d'entreprise	942	1 357
Techniciens	678	756
Total population active	23 172	24 748

Source : INSEE, *Données sociales*, 1990.

On constate que l'effectif des cadres d'entreprise a augmenté. Il en va de même des techniciens.

Si l'on s'intéresse à la structure sociale et à son évolution, c'est en proportions qu'il faut raisonner. De 1982 à 1990, en effet, l'effectif de la population active *dans son ensemble* a augmenté : la croissance d'une catégorie pourrait être entièrement imputable à la croissance de la population active en général. Dans cette optique, on comparera non pas les effectifs mais, par exemple, leur traduction en pour cent :

— cadres en 1982 : 100 × 942/23172 = 4,07 ;
— cadres en 1990 : 100 × 1357/24748 = 5,48.

On aura rapporté 942 à 23172 et 1357 à 24748.

1. J.-M. Faverge, *Méthodes statistiques en psychologie appliquée*, t. III, chap. XVI, Paris, PUF, 1965 ; J.-P. Benzécri *et al.*, *L'analyse des données*, t. IB, nº 5, Paris, Dunod, 1973 ; M. Volle, *Analyse des données*, chap. III, Paris, Economica, 1985.

Calculer des pourcentages, c'est raisonner toutes choses égales par ailleurs en ce qui concerne les effectifs totaux. En les ramenant à 100, c'est faire comme si ces effectifs totaux n'avaient pas varié.

Les cadres sont donc en réelle progression. La progression des techniciens dans la structure des actifs est moins nette : pour des effectifs de 678 000 en 1982 et de 756 000 en 1990, leur part en pour cent dans la population active n'est passée que de 2,93 à 3,05.

3. **Des cas très particuliers.** — Il ressort de la règle précédente que raisonner à effectifs totaux égaux (ramenés à 100) c'est *admettre que les proportions ne dépendent pas de l'effectif total.* Le sondage ou l'expérimentation sont des cas typiques où cette hypothèse est valide : que l'on sonde ou expérimente sur 2 000 personnes au lieu de 1 000 ne changera pratiquement rien sur les proportions observées de « satisfaits » ou de « guéris ». Il est cependant des cas où cette hypothèse est illicite et où, par conséquent, le calcul de pourcentages est dangereux. Donnons deux exemples de situations invalidantes, où les proportions sont, d'une certaine manière, déterminées par l'effectif total.

A) *Cas d'une détermination structurelle.* — On appelle étendue du vocabulaire d'un texte le nombre V de vocables (ou mots différents, pour simplifier) qu'il contient. Chacun de ces vocables, ou bien n'apparaît qu'une seule fois dans le texte (c'est un *hapax*), ou bien apparaît plusieurs fois. V_1 est le nombre d'hapax. $V - V_1$ est le nombre de vocables qui se répètent au moins une fois. $(V - V_1)/V$ est le « taux de répétition ». Cette proportion peut-elle servir à comparer le style de deux textes sous l'angle de la richesse du lexique ? La réponse est négative car le taux de répétition dépend de l'étendue du vocabulaire. Dans le tableau 3, on a pris un texte de 656 vocables chez Julien Gracq et un texte, plus court, de 369 vocables chez Louis Aragon :

Tableau 3. — **Taux de répétition**

	Gracq	*Aragon*
Hapax	427	263
Autres	229	106
Total	656	369

Source : Ch. Muller, *Principes et méthodes de statistique lexicale*, Paris, Hachette, 1977.

En comparant les taux de répétition 229/656 =35 % et 106/369 = 29 %, il semblerait que le style de Gracq soit plus redondant que celui d'Aragon. Mais en fait, au lieu de comparer le style des deux auteurs, on n'a fait que vérifier un phénomène général : plus on est long et plus on se répète. Pour comparer les styles, il aurait fallu prendre des textes de longueurs identiques !

B) *Cas d'une détermination statistique.* — C'est le cas où les effectifs sont très faibles de sorte que les proportions sont soumises à de très fortes variations. Supposons que nous ayons réalisé quelques entretiens pour explorer les motivations des étudiants à l'entrée à l'université. Jusqu'à quel niveau, par exemple, envisagent-ils de poursuivre leurs études ? On obtient, sur 10 étudiants de psychologie et 12 de sociologie :

Tableau 4. — **Un mini-sondage**

	Psychologie	*Sociologie*
Thèse	1	2
Autre réponse	9	10
Total	10	12

Peut-on rapporter 1 à 10 et 2 à 12 en s'exprimant par des pourcentages (10,0 % et 16,7 %) ? Certes non, car le déplacement d'un seul individu modifierait du tout au tout les proportions et les différences entre ces proportions (1/10 pourrait devenir 0/10 soit 0 % et 2/12 devenir 3/12 soit 25 % : l'écart entre les pourcentages passerait de 6,7 à 25 % !). Une expression en pourcentages serait donc bien malvenue. Elle laisserait entendre qu'on prétend estimer avec précision les proportions qu'on observerait avec de plus grands échantillons ou par recensement des populations concernées.

III. — Fréquences, taux, probabilités...

Fréquences, proportions, pourcentages, taux, ratios, rapports, quotients, points, indices, probabilités, chances : autant de termes qui montrent la diversité du vocabulaire lié au fait de diviser pour rapprocher. Il semble utile d'en examiner les définitions et les usages. Nous partirons de l'exemple suivant (tableau 5) où on a dénombré les effectifs masculins et féminins dans la catégorie des « cadres et professions intellectuelles supérieures » en 1962 et en 1989.

Tableau 5. — **Les diverses possibilités de rapprochement**

	1962	*1989*
Hommes	747	1 650
Femmes	**145**	**668**
Total	892	2 318
Proportion de femmes	0,163	0,288
Pourcentage de femmes	16,3 %	28,8 %
Femmes, en pour cent	16,3	28,8
Ratio femmes/hommes	0,194	0,405
Variation relative de l'effectif femmes		
En proportion		3,607
En pourcentage		360,7 %
En pour cent		360,7
Indice base 100	100	461
Coefficient multiplicateur		4,607

Les effectifs sont donnés en milliers. Lire : 747 000 hommes cadres en 1962, etc.

D'après A. Chenu, Les groupes socioprofessionnels, *Nouveau Manuel. Sciences économiques et sociales*, Paris, La Découverte, 1995.

1. **Proportions et pourcentages.** — Étudions le groupe professionnel en 1962. On a 747 pour les hommes et 145 pour les femmes, sur un total de 892.

A) *Conventions d'écriture.* — La proportion des hommes est 747/892 = 0,837, celle des femmes est 145/892 = 0,163. Ce qui distingue le calcul d'une proportion d'autres types de divisions est que le numérateur de la fraction *affecte* le dénominateur. Les 747(000) hommes du numérateur font partie des 892(000) personnes du dénominateur : 0,837 = 747/(747 + 145). C'est la raison pour laquelle la somme des proportions d'hommes (0,837) et de femmes (0,163) est égale à l'unité.

L'emploi courant du mot « proportion » a évolué. On appelait encore au début du siècle « proportion » non pas un nombre (une valeur numérique), mais le fait de l'égalité entre des rapports : « 3/4 = 6/8 » est une proportion au sens d'une relation de proportionnalité. Mais on a pris l'habitude de nommer directement « proportion » le nombre qui représente la classe d'équivalence des fractions égales entre elles : 3/4 = 6/8 = 1,5/2 = . . . = 0,75/1 = 0,75 (numérateur de la fraction dont le dénominateur est égal à 1). C'est la convention que nous utiliserons.

Il convient également d'être clair sur la convention d'écriture utilisée lorsqu'on emploie le symbole « x % » (pour cent). Quand nous écrivons que les femmes représentent 16,3 % des cadres en 1962, nous pouvons interpréter ce pourcentage de deux manières, descriptive ou opératoire.

Sur le plan descriptif, le pourcentage montre comment se répartit une population indépendamment de son effectif : pour 100 cadres, 16,3 sont des femmes. En ce sens, la somme de 16,3 et de 83,7 est égale à 100 unités.

Sur un autre plan, le pourcentage est une *opération* qui consiste à multiplier par 16,3 et à diviser par 100 (ou à multiplier par 0,163, ce qui revient au même). On peut alors :

— reconstituer (avec une approximation due aux arrondis) les effectifs partiels quand on connaît les effectifs totaux : 0,163 × 892 = 145,4 ;

— calculer les effectifs qu'on aurait dû obtenir sur une autre population si les proportions étaient conservées : $0{,}163 \times 2318 = 378$ milliers de femmes qu'on aurait dû avoir en 1989.

La convention qui en découle est d'écrire 16,3 dans le cas descriptif et 16,3 % ou 0,163 dans le cas opératoire, utilisant alors indifféremment le terme de pourcentage ou de proportion.

B) *La division par cent.* — Pour une opération sur calculette, il est clair qu'il est préférable d'utiliser la proportion que le pour cent. Multiplier par 16,3 et diviser par 100 exige en effet deux opérations ; multiplier par 0,163 n'en demande qu'une seule !

En revanche, le pour cent reflète mieux le raisonnement classique de la règle de trois : 145 est à 892 comme 16,3 est à 100.

C) *Base d'un pourcentage.* — On appelle « base » du calcul de proportion ou de pourcentage le dénominateur de la fraction concernée, c'est-à-dire 892 dans l'exemple. La base est l'effectif total auquel on rapporte tout ou partie des effectifs qui le composent.

A partir de quelle base minimum peut-on présenter un pourcentage ? Des effectifs minima comme 50 ou 30 sont parfois cités. Une règle cohérente serait d'exiger que l'effectif total soit au moins égal à 100 puisque les effectifs partiels sont « ramenés à 100 ». Mais avec un tel principe, bien des enquêtes s'effondreraient ! En fait, le problème de la signification statistique d'un pourcentage se résout par le calcul des probabilités, dans le cadre du modèle de l'échantillonnage. On trouvera dans tout manuel de statistique inductive les méthodes de détermination de la précision d'un pourcentage. On montre, par exemple, qu'un pourcentage de 50 % calculé sur un échantillon de $N = 1\,000$ personnes (effectif de la plupart des sondages) est affecté d'une erreur de ± 3 % (pour simplifier, utiliser la formule $\pm 1/\sqrt{N}$). Ainsi, dire que la cote d'un homme politique est passée de 49 % à 51 % a peu de sens, étant donnée l'imprécision sur chacune des valeurs. Dans un tel cadre inductif, la valeur minimum de la base peut être ainsi déterminée par le niveau de précision souhaité.

2. **Fréquence.** — Ce terme est fréquemment utilisé, mais de manière très variable ! Bien des auteurs désignent par « fréquence » aussi bien les grandeurs absolues (effectifs) que les grandeurs relatives (proportions).

On trouvera même les expressions « fréquences absolues » et « fréquences relatives » pour distinguer effectifs et proportions ! Dans la tradition linguistique on dira par exemple que dans la phrase précédente la fréquence de la lettre *e* est 20 (fréquence absolue) ou 20/129 (fréquence relative).

Le mot « fréquence » renvoie à ce qui survient souvent, pas souvent, à ce qui est attendu, inattendu, etc. Il suggère les notions de temps et de probabilité, notions qui n'ont souvent aucun lien avec le phénomène étudié.

Dans le domaine de l'induction statistique on oppose « fréquence » (proportion observée sur l'échantillon) à « probabilité » (proportion dans la population), le problème étant de produire un jugement sur la population à partir de l'observation d'un échantillon. Nous limitant au domaine de la statistique descriptive, et pour des raisons de polysémie, nous préférerons éviter ce terme de fréquence dans la suite de cet ouvrage.

3. **Points.** — « M. X... a perdu 10 points de popularité », « un point de croissance a été gagné » : voilà des expressions familières. De 1962 à 1989, le pourcentage de femmes passe dans notre exemple de 16,26 % à 28,82 %. La différence est de « 12,56 % ». Mais cette valeur est-elle un pourcentage ? La réponse est oui, si l'on considère que sur 2 318 personnes en 1989 on a $12{,}56 \times 2\,318/100 = 291$ femmes de plus que ne l'aurait laissé espérer le maintien du pourcentage initial de 16,26 %. En revanche, dire que le pourcentage « a augmenté de 12,56 % » pourrait laisser entendre que le nouveau pourcentage est de $16{,}26 + (12{,}56\,\% \times 16{,}26) = 18{,}30\,\%$, ce qui est évidemment faux. 12,56 % n'est pas une variation relative mais un écart de pourcentages. Pour éviter cette ambiguïté, on s'exprime en « points » : le pourcentage de femmes parmi les cadres a gagné 12,56 points entre les deux époques. Il faut toutefois

reconnaître que ce vocabulaire évoque une idée de performance (politique, économique...) que l'on ne tient pas forcément à connoter.

4. **Ratios, rapports.** — Lorsque le rapport d'effectifs ne comporte pas la base en dénominateur mais consiste en un rapport de deux effectifs partiels, il ne s'agit plus de proportion ou de pourcentage. Selon le domaine et les habitudes, on parle de ratio, de rapport, de coefficient...

Dans l'exemple du tableau 5, le ratio femmes/hommes est égal à 145/747 = 0,194. On parlera d'un rapport de féminité égal à 0,194.

Les démographes utilisent plutôt le rapport de masculinité qui vaudrait 747/145 = 5,152. Le rapport de masculinité à la naissance est égal à 1,05 environ (proportionnalité de 105 naissances masculines pour 100 naissances féminines, découverte en Angleterre par John Graunt, vers 1660).

On parlera également du ratio vote de droite / vote de gauche, ou du « rapport démographique » : nombre d'actifs occupés/nombre d'inactifs retraités, ou encore du « rapport de dépendance » : (les 0-14 ans + les 65 ans et plus) / (les 15-64 ans), etc.

De manière générale, un ratio ou un rapport est une fraction dont les termes sont homogènes. Ces termes sont de même nature et figurent dans une somme totale. Un ratio rapproche donc deux effectifs homogènes particulièrement intéressants.

5. **Taux, quotients, coefficients.** — L'appellation « taux » est très peu contrôlée mais très employée. On en dénombre des dizaines d'occurrences dans tout bon dictionnaire économique[1] mais elle signifie tout aussi bien pourcentage ou ratio que variation relative. Le mot peut aussi laisser entendre qu'on a calculé une fraction où la relation logique entre numérateur N et dénominateur D est plus ou moins rigoureuse. Donnons quelques exemples empruntés à la démographie[2] ou à l'économie.

1. Par exemple : J.-Y. Capul, O. Garnier, *Dictionnaire d'économie et de sciences sociales*, Paris, Hatier, 1994.
2. R. Pressat, *Démographie statistique*, Paris, PUF, 1980.

A) *La relation logique entre N et D est lâche.* — Le taux brut de mortalité, pour une année donnée, est le rapport entre le nombre de décès et l'effectif de la population pendant l'année :

Taux brut de mortalité = N/D
N = nombre de décès dans l'année
D = population moyenne de l'année.

Comme le souligne Roland Pressat, un tel taux n'est pas une proportion mais une *fréquence annuelle d'événements au sein d'une population,* estimée de manière plus ou moins arbitraire.

B) *La relation logique entre N et D est serrée.* — A l'inverse du taux brut de mortalité le quotient de mortalité, est une réelle proportion et, à ce titre, une estimation de probabilité de décès :

Quotient de mortalité à l'âge x = N/D
N = décès entre les âges x et $x + 1$
D = survivants à l'âge x.

Alors que le terme « quotient » marque la différence avec un taux qui ne serait pas une proportion et encore moins un « risque », il y a de nombreux cas où des proportions sont appelées taux. Le taux de femmes parmi les cadres en 1962 serait dans notre exemple de 16,3 %. On parlera ainsi de taux de chômage :

Taux de chômage = N/D
N = chômeurs
D = chômeurs + actifs occupés.

Remarquons que ce taux est une proportion mais qu'il n'est pas pour autant un risque : les fonctionnaires, qui figurent en dénominateur, sont en effet prémunis du chômage.

C) *La relation logique entre N et D est complexe ou peu définie.* — On préfère alors une formulation qui explicite les deux termes de la fraction :

Le PIB par tête = N/D
N = produit intérieur brut
D = population totale.

Au total, il y a (au moins) trois raisons diverses pour parler de taux :

- la relation entre numérateur et dénominateur n'est pas nettement celle d'un rapport de proportionnalité (non présence du numérateur dans le dénominateur) ; le terme de taux évoque cette *hétérogénéité* ;
- une catégorie de la variable nominale est *distinguée* des autres et devient le centre d'intérêt : on parlera par exemple du taux de célibataires dans une classe d'âge ;
- c'est d'une *variation* dont il est question, et c'est ce que nous allons considérer maintenant.

6. **Variations relatives.** — Revenons au tableau 5 pour explorer le second type d'applications du calcul des proportions : la mesure des variations relatives. L'effectif des femmes est passé de 145 en 1962 à 668 en 1989. L'accroissement de population est donc de 668 – 145 = 523. Rapporté à l'effectif initial 145, l'accroissement relatif, ou variation relative, est en pour cent 523/145 × 100 = 360,7 ou 3,607 en proportion. En économie, on exprime souvent une variation relative en ramenant la valeur initiale x à une base 100. Ce qui donnerait ici, après arrondi, 461 pour la valeur finale :

$$x/100 = 668/145 \text{ donne } x = 100 \times 668/145 = 460{,}7.$$

Bien entendu, le calcul de variation relative et le calcul de l'indice base 100 sont équivalents : on retrouve 361 dans l'indice 461 = 100 + 361. Il est encore équivalent de dire qu'on a multiplié la valeur initiale par le « coefficient multiplicateur » (1 + 3,607) = 4,607.

Notons que l'évolution des femmes aurait pu tout aussi bien être chiffrée par la variation relative du *taux* de femmes :

$$v = (28{,}8 - 16{,}3)/16{,}3 = 0{,}767 \text{ ou } 76{,}7\ \%.$$

Leur poids dans la profession a augmenté de 76,7 % entre les deux périodes, variation relative qui n'a rien à voir avec la précédente et qui mesure l'évolution de la *structure* de la profession.

Formellement, soient Qd (départ) et Qa (arrivée) deux quantités ($Qd > 0$, $Qa \geqslant 0$). On a :

- variation relative en proportion : $v = (Qa - Qd)/Qd$
- variation relative en pour cent : $V = 100 \times (Qa - Qd)/Qd$
- coefficient multiplicateur : $C = Qa/Qd$
- indice base 100 : $I = 100 \times Qa/Qd$

ainsi que les relations :

$V = 100v \quad I = 100C \quad C = 1 + v \quad I = 100 + V$

v (ou V) s'appellent aussi taux de croissance ou taux d'accroissement. Ces quantités peuvent être négatives (si $Qa < Qd$) ou dépasser 1 ou 100 (si $Qa > 2Qd$), comme dans l'exemple. Il ne faut donc pas confondre variation relative et valeur relative ! De très nombreux taux, en économie, sont des variations relatives : taux de croissance, taux d'inflation, taux d'intérêt, taux de TVA, etc. Ce dernier taux mérite bien son nom puisqu'il est une taxe. S'il est de 18,6 %, un produit qui vaut hors taxe 100 F (prix de départ HT) sera payé toutes taxes comprises 100 + 18,6 = 118,6 F (prix d'arrivée TTC). On a : 18,6 = (118,6 – 100)/100.

Le taux de TVA = (TTC – HT)/HT est bien une variation relative. Et rien ne l'empêche d'évoluer dans le temps et de gagner... 2 points en passant à 20,6 en 1995 ! Remarquons enfin que le calcul pratique du prix TTC se fait au moyen du coefficient multiplicateur : TTC = HT × 1,206.

7. **Probabilités.** — Revenons aux effectifs exprimés en valeur relative. On est souvent tenté d'exprimer ces valeurs relatives (proportions ou pourcentages) en termes de probabilités, de chances, de risques. Cette opération de la pensée est justifiée dans certains domaines (physique, biologie...) ou pour certaines applications (assurances, santé...). Mais dans d'autres domaines la lecture probabiliste des pourcentages sera génératrice d'erreurs.

Prenons les élèves entrés en 6^{e} en 1980[1]. 46 % de ces élèves accèdent à la seconde. Peut-on dire que, pour chacun d'entre eux pris le jour de leur entrée en 6^{e}, la probabilité était de 46 chances sur 100 d'entrer un jour en seconde ? Une telle proposition est absurde. Preuve en est que le taux d'accès en

1. INSEE, *Données sociales*, 1993.

seconde s'élève à 90 % pour les enfants d'enseignants et n'est que de 26 % pour les enfants des catégories les plus modestes. La probabilité individuelle ne saurait varier selon l'information dont on dispose sur la personne !

En réalité, 46 % représente effectivement une probabilité, mais c'est une probabilité qui est attachée à l'observateur et non pas à chaque enfant. Si *je* me dispose à prendre au hasard, en 1995, un de ces enfants entrés en 6e en 1980, *j*'ai effectivement 46 chances sur 100 de constater qu'il a accédé à la seconde. Si *je* restreint l'ensemble de définition aux enfants d'enseignants, *ma* probabilité devient 90 %.

Ce principe pratique ou « praxéologique » qui associe l'observateur aux observations est tout à fait rigoureux et diffère radicalement du passage illicite – que l'on effectue souvent avec discrétion – d'une caractéristique collective à une caractéristique individuelle.