

Statistiques descriptives

Cours 2

Structure des variables

Structure des variables

- ◆ Considérons la définition suivante
« Une variable est une propriété d'un objet qui peut prendre différentes valeurs. Ainsi la couleur des cheveux est une variable parce qu'il s'agit de la propriété d'un objet (les cheveux) et qu'elle peut prendre différentes valeurs (cheveux bruns, blonds, roux, gris, etc.) »
David C. Howell (2008, p. 4)
- ◆ On voit qu'ainsi définie la notion de variable est une notion extrêmement générale, qui nous permettra de décrire des objets sous l'angle des valeurs que peuvent prendre les propriétés de ces objets.

Structure des variables

- ◆ Variable (ou caractère) statistique, valeurs
 - ✓ Une variable est une information dont on recueille (ou observe ou mesure) la valeur sur *chaque* individu. On parle de variable parce que la valeur de l'information n'est pas la même d'un individu à l'autre.
- ◆ Il existe 2 types de variables : qualitatives & quantitatives

Structure des variables

◆ Les variables qualitatives

- ✓ Elles s'expriment en modalités. Les modalités sont comme des choix de réponses aux variables étudiées.

◆ Les variables quantitatives

- ✓ Sont quant à elles des variables représentées par des quantités telles que l'âge, le poids et la taille. Elles s'expriment en valeurs. Les valeurs représentent les choix de réponses aux variables quantitatives.

Structure des variables

- ◆ Exemple : soit un test de Q.I on s'intéresse à 2 sujets, disons le sujet 12 et 18

Sujets	Âges	Professions	...
...
12	40	4	...
18	38	2	...
...

- ◆ La différence d'âge ($40-38=2$) peut-elle être traitée de la même manière que la différence de profession du patient ($4-2=2$) ?
- ◆ On dira que l'échelle des âges et l'échelle des professions ont des caractéristiques différentes

Structure des variables

- ◆ **Lorsqu'on fait une mesure en psychologique**
 - ✓ Il est nécessaire de savoir comment on va procéder à cette mesure
 - ✓ On utilise toujours une échelle de mesure pour reporter la valeur de ce que l'on observe
 - ✓ Attention : mesurer n'implique pas obligatoirement l'utilisation du système numérique.

- ◆ **Catégorisation d'échelles numériques dites « de mesure ».**
On rencontre quatre d'échelles:
 - ✓ l'échelle nominale,
 - ✓ l'échelle ordinale,
 - ✓ échelle intervalle (relative),
 - ✓ l'échelle rapport (absolue).

S. S. Stevens publie en 1946 son célèbre article « On the theory of scales of measurement* » contribution majeure du 20e siècle dans le domaine de la mesure en psychologie

SCIENCE

Vol. 103, No. 2684

Friday, June 7, 1946

678

SCIENCE

Vol. 103, No. 2684

On the Theory of Scales of Measurement

S. S. Stevens

Director, Psycho-Acoustic Laboratory, Harvard University

FOR SEVEN YEARS A COMMITTEE of the British Association for the Advancement of Science debated the problem of measurement. Appointed in 1932 to represent Section A (Mathematical and Physical Sciences) and Section J (Psychology), the committee was instructed to consider and report upon the possibility of "quantitative estimates of sensory events"—meaning simply: Is it possible to measure human sensation? Deliberation led only to disagreement, mainly about what is meant by the term measurement. An interim report in 1938 found one member complaining that his colleagues "came out by that same door as they went in," and in order to have another try at agreement, the committee begged to be continued for another year.

For its final report (1940) the committee chose a common bone for its contentions, directing its arguments at a concrete example of a sensory scale. This was the Sone scale of loudness (S. S. Stevens and H. Davis. *Hearing*. New York: Wiley, 1938), which purports to measure the subjective magnitude of an auditory sensation against a scale having the formal properties of other basic scales, such as those used to measure length and weight. Again the 19 members of the committee came out by the routes they entered, and their views ranged widely between two extremes. One member submitted "that any law purporting to express a quantitative relation between sensation intensity and stimulus intensity is not merely false but is in fact meaningless unless and until a meaning can be given to the concept of addition as applied to sensation" (Final Report, p. 245).

It is plain from this and from other statements by the committee that the real issue is the meaning of measurement. This, to be sure, is a semantic issue, but one susceptible of orderly discussion. Perhaps agreement can better be achieved if we recognize that measurement exists in a variety of forms and that scales of measurement fall into certain definite classes. These classes are determined both by the empirical operations invoked in the process of "measuring" and

by the formal (mathematical) properties of the scales. Furthermore—and this is of great concern to several of the sciences—the statistical manipulations that can legitimately be applied to empirical data depend upon the type of scale against which the data are ordered.

A CLASSIFICATION OF SCALES OF MEASUREMENT

Paraphrasing N. R. Campbell (Final Report, p. 340), we may say that measurement, in the broadest sense, is defined as the assignment of numerals to objects or events according to rules. The fact that numerals can be assigned under different rules leads to different kinds of scales and different kinds of measurement. The problem then becomes that of making explicit (a) the various rules for the assignment of numerals, (b) the mathematical properties (or group structure) of the resulting scales, and (c) the statistical operations applicable to measurements made with each type of scale.

Scales are possible in the first place only because there is a certain isomorphism between what we can do with the aspects of objects and the properties of the numeral series. In dealing with the aspects of objects we invoke empirical operations for determining equality (classifying), for rank-ordering, and for determining when differences and when ratios between the aspects of objects are equal. The conventional series of numerals yields to analogous operations: We can identify the members of a numeral series and classify them. We know their order as given by convention. We can determine equal differences, as $8 - 6 = 4 - 2$, and equal ratios, as $8/4 = 6/3$. The isomorphism between these properties of the numeral series and certain empirical operations which we perform with objects permits the use of the series as a *model* to represent aspects of the empirical world.

The type of scale achieved depends upon the character of the basic empirical operations performed. These operations are limited ordinarily by the nature of the thing being scaled and by our choice of procedures, but, once selected, the operations determine

that there will eventuate one or another of the scales listed in Table 1.²

The decision to discard the scale names commonly encountered in writings on measurement is based on the ambiguity of such terms as "intensive" and "extensive." Both ordinal and interval scales have at

Thus, the case that stands at the median (mid-point) of a distribution maintains its position under all transformations which preserve order (isotonic group), but an item located at the mean remains at the mean only under transformations as restricted as those of the linear group. The ratio expressed by the coefficient

TABLE 1

Scale	Basic Empirical Operations	Mathematical Group Structure	Permissible Statistics (invariantive)
NOMINAL	Determination of equality	Permutation group $x' = f(x)$ $f(x)$ means any one-to-one substitution	Number of cases Mode Contingency correlation
ORDINAL	Determination of greater or less	Isotonic group $x' = f(x)$ $f(x)$ means any monotonic increasing function	Median Percentiles
INTERVAL	Determination of equality of intervals or differences	General linear group $x' = ax + b$	Mean Standard deviation Rank-order correlation Product-moment correlation
RATIO	Determination of equality of ratios	Similarity group $x' = ax$	Coefficient of variation

times been called intensive, and both interval and ratio scales have sometimes been labeled extensive.

It will be noted that the column listing the basic operations needed to create each type of scale is cumulative: to an operation listed opposite a particular scale must be added all those operations preceding it. Thus, an interval scale can be erected only provided we have an operation for determining equality of intervals, for determining greater or less, and for determining equality (not greater and not less). To these operations must be added a method for ascertaining equality of ratios if a ratio scale is to be achieved.

In the column which records the group structure of each scale are listed the mathematical transformations which leave the scale-form invariant. Thus, any numeral, x , on a scale can be replaced by another numeral, x' , where x' is the function of x listed in this column. Each mathematical group in the column is contained in the group immediately above it.

The last column presents examples of the type of statistical operations appropriate to each scale. This column is cumulative in that all statistics listed are admissible for data scaled against a ratio scale. The criterion for the appropriateness of a statistic is *invariance* under the transformations in Column 3.

²A classification essentially equivalent to that contained in this table was presented before the International Congress for the Unity of Science, September 1941. The writer is indebted to the late Prof. G. D. Birkhoff for a stimulating discussion which led to the completion of the table in essentially its present form.

of variation remains invariant only under the similarity transformation (multiplication by a constant). (The rank-order correlation coefficient is usually deemed appropriate to an ordinal scale, but actually this statistic assumes equal intervals between successive ranks and therefore calls for an interval scale.)

Let us now consider each scale in turn.

NOMINAL SCALE

The *nominal scale* represents the most unrestricted assignment of numerals. The numerals are used only as labels or type numbers, and words or letters would serve as well. Two types of nominal assignments are sometimes distinguished, as illustrated (a) by the 'numbering' of football players for the identification of the individuals, and (b) by the 'numbering' of types or classes, where each member of a class is assigned the same numeral. Actually, the first is a special case of the second, for when we label our football players we are dealing with unit classes of one member each. Since the purpose is just as well served when any two designating numerals are interchanged, this scale form remains invariant under the general substitution or permutation group (sometimes called the symmetric group of transformations). The only statistic relevant to nominal scales of Type A is the number of cases, e.g. the number of players assigned numerals. But once classes containing several individuals have

- ◆ Stevens est le fondateur d'une branche de la psychophysique contemporaine prônant la mesure directe des sensations. Il a enseigné la psychophysiologie et la psychologie à Harvard à partir de 1938 et y a dirigé le laboratoire de psychophysique



- ◆ (Wiki) Il contribua à la psychophysique en proposant une loi de puissance reliant la grandeur physique d'un stimulus et l'intensité perçue, dite loi de Stevens.

Structure des variables

◆ Les variables qualitatives expriment une qualité. On parlera:

✓ D'échelle nominale

Une variable est dite qualitative nominale quand ses valeurs sont des éléments d'une catégorie non hiérarchique. En d'autres termes, ses éléments ne peuvent pas se ranger dans une gradation logique, selon une hiérarchie naturelle (il n'existe aucune relation d'ordre).
Exemple: la *couleur*, situation familiale (Marié, Pacsé ...)

✓ D'échelle ordinale

Une variable qualitative ordinale possède toutes les propriétés de la variable qualitative nominale avec en plus la possibilité de positionner et de hiérarchiser les individus entre eux selon la valeur attachée à leur caractère.

Exemple: échelle de Likert* (1: pas du tout d'accord, 2 ... 5 : tout à fait d'accord)

* Échelle de jugement répandue dans les questionnaires psychométriques par laquelle la personne interrogée exprime son degré d'accord ou de désaccord vis-à-vis d'une affirmation.

Structure des variables

◆ Pour les variables quantitatives

- ✓ D'intervalle (relative), lorsque l'on peut calculer l'écart entre deux valeurs d'une variable quantitative, la division et la multiplication n'ont pas de sens.

La valeur «0» a été fixée par convention

- ✓ Certaines opération arithmétiques sont pourvues de sens mais d'autres restent délicates car leur résultat ne correspond pas à une réalité

Soit deux valeurs L_1 et L_2 :

- On peut dire L_2 à x de plus (de moins) que L_1
- On ne peut pas dire (pas de sens) que L_2 est x fois plus grand (plus petit) que L_1

Structure des variables

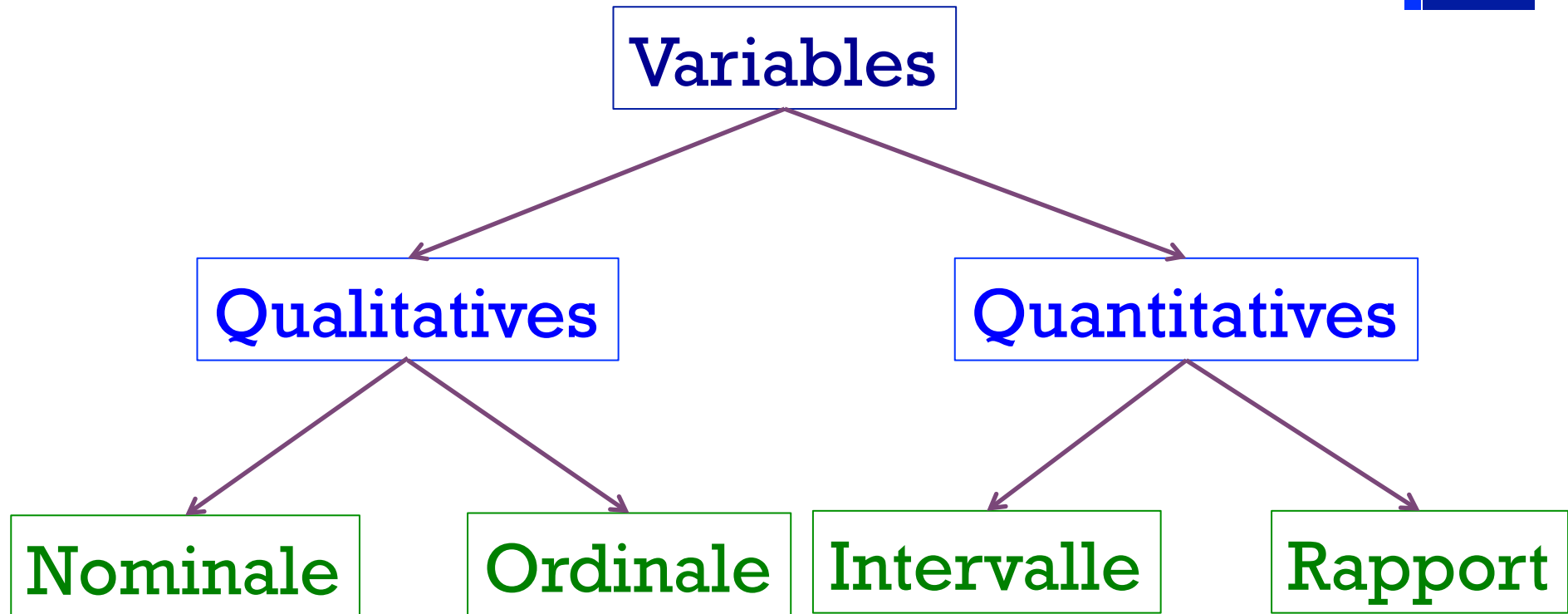
◆ Pour les variables quantitatives

- ✓ De rapport ou de ratio (absolue), lorsqu'on peut calculer un rapport entre deux valeurs (en plus de l'intervalle).
Le zéro signifie une absence de caractère.
- ✓ Elle est peu utilisées en psychologie. Il s'agit d'un cas particulier de mesure d'intervalles.
- ✓ Les échelles de rapport présentent les propriétés numériques les plus élaborées et qui permettent toutes les opérations arithmétiques en leur conservant une signification.
 - On peut dire L_2 à x de plus (de moins) que L_1
 - On peut dire que L_2 est x fois plus grand (plus petit) que L_1

Structure des variables

Variables	Qualitatives (ne prend pas de valeurs numériques)	Quantitatives ou numériques (continues ou discrètes)
Echelles	<ul style="list-style-type: none">- Nominales Pas d'ordre naturel- Ordinales ordre naturel	<ul style="list-style-type: none">- D'intervalles: zéro situé de manière arbitraire, un zéro ne correspond pas à une absence- De rapport: zéro signification précise absence du caractère considère

En résumé



Quelles sont les natures des variables et les échelles utilisées pour ces différentes mesures

Un questionnaire porte sur:

- ◆ L'âge
- ◆ La température de l'eau (°C)
- ◆ Echelle d'évaluation du cours de stats.
J'aime : pas du tout, un peu, ... , énormément
- ◆ La taille
- ◆ Le nombre d'enfants d'une personne
- ◆ La situation familiale (Marié, Pacsé...)

Quelles sont les natures des variables et les échelles utilisées pour ces différentes mesures

Un questionnaire porte sur:

- ◆ L'âge: **quantitative continue, échelle de rapports**
- ◆ La température de l'eau (°C): **quantitative continue, échelle d'intervalle**
- ◆ Echelle d'évaluation du cours de stats.
J'aime : pas du tout, un peu, ... , énormément: **qualitative, échelle ordinale**
- ◆ Le nombre d'enfants d'une personne: **quantitative discrète, échelle de rapports**
- ◆ La situation familiale (Marié, Pacsé...): **qualitative, échelle nominale**

Série, effectif, fréquence et tableau statistique

- ◆ Si des observations, donc des mesures, sont répétées un nombre de fois important, on obtient un ensemble de valeurs de la variable caractérisé par son effectif total, c'est à dire le nombre d'observations, lequel se répartit en effectifs partiels.
- ◆ Il est habituel en statistique de désigner par le symbole x les valeurs de la variable. Par exemple, x_4 constituera la 4eme valeur observée, x_{127} la 127eme valeur, etc...
- ◆ Il ne faut pas confondre le x en algèbre, qui désigne une grandeur dont on cherche à connaître la valeur avec le x statistique qui désigne une grandeur observée donc connue.

Série, effectif, fréquence et tableau statistique

- ◆ On appelle série statistique la suite des valeurs prises par une variable X sur les unités d'observation. Le nombre d'unités d'observation est noté n . Les valeurs de la variable X sont notées :

$$x_1, x_2, \dots, x_k, \dots, x_{n-1}, x_n$$

- ◆ Un tableau statistique est une méthode permettant de visualiser des données. Il permet une meilleure compréhension des résultats de recherches statistiques en regroupant des informations de même nature, regroupement par classes.
- ◆ L'effectif d'une classe est le nombre d'éléments de la population observés dans cette classe, n est l'effectif de l'échantillon.

Série, effectif, fréquence et tableau statistique

- ◆ Fréquence absolue : effectif des observations d'une classe
- ◆ Fréquence relative (ou simplement fréquence)
Quotient de cet effectif par celui de la population.
 - ✓ Si valeur est un nombre entier positif, il s'agit de la fréquence absolue, c'est-à-dire l'effectif de la classe.
 - ✓ Si valeur est un nombre compris entre 0 et 1 ou un pourcentage, il s'agit de la fréquence relative.
- ◆ Aide pour les comparaisons entre des séries d'observations portant sur des populations inégalement nombreuses.
- ◆ L'expression en pourcentage (%) facilite les comparaisons

Série, effectif, fréquence et tableau statistique

◆ Exemple

Répartition des 30 élèves d'une classe de 4^e selon leur âge à la rentrée.

Ages	13	14
Effectifs	18	12
Fréquences	0.6	0.4



Ou fréquence absolue

Série, effectif, fréquence et tableau statistique

◆ Exemple

On connaît les QI obtenus à un test par 72 sujets, par exemple 72 enfants de 10 à 11 ans. Les résultats s'étendent de 81 à 127 (intervalle de variation).

113	109	101	82	86	102	104	98	114	106
97	111	121	96	103	93	108	112	111	109
99	94	117	90	92	106	127	105	100	101
119	97	115	85	107	89	116	102	104	103
99	100	98	100	91	120	110	105	96	95
113	118	110	118	81	110	93	102	107	103
97	111	123	101	106	105	108	107	117	95
88	103								

Série, effectif, fréquence et tableau statistique

◆ Exemple

On connaît les QI obtenus à un test par 72 sujets, par exemple 72 enfants de 10 à 11 ans. Les résultats s'étendent de 81 à 127 (intervalle de variation).

- ◆ On peut soupçonner que cet échantillon n'est pas représentatif de l'ensemble de tous les enfants de 10 à 11 ans. Il n'est sans doute pas tiré au hasard puisqu'il n'y a pas de $QI < 81$ ou $QI > 127$ (on peut avoir des $QI < 20$ ou $QI > 140$)

Série, effectif, fréquence et tableau statistique

◆ Exemple

On connaît les QI obtenus à un test par 72 sujets, par exemple 72 enfants de 10 à 11 ans. Les résultats s'étendent de 81 à 127 (intervalle de variation).

◆ Groupe par classe

- ✓ Relevé et rangé
- ✓ En 10 classes
- ✓ Effectifs tot. $N = \sum n_k = 72$
- ✓ $\sum (\text{fréquence})_k = 100\%$

QI	Effectifs (n_k)	Fréquences (%)
82	2	2,78
87	4	5,55
92	6	8,33
97	11	15,27
102	15	20,83
107	13	18,06
112	10	13,89
117	7	9,73
122	3	4,17
127	1	1,39

Statistiques descriptives

- ◆ Un premier objectif des statistiques descriptives est de décrire au mieux les échantillons, c'est-à-dire de les synthétiser au travers d'indices descriptifs, appelés également résumés statistiques. Ces résumés visent avant tout à donner des renseignements sur la tendance centrale et sur la dispersion de la distribution
- ◆ On va parler dans ce qui suit de:
 - ✓ Mode
 - ✓ Moyenne
 - ✓ Médiane
 - ✓ Mesure de dispersion