

Description des données en statistique et biostatistique : Types de variables

Description of data in statistics and biostatistics: Types of variables

Pefura-Yone Eric Walter^{1,2,*}, Djenabou Amadou^{2,3}, Balkissou Adamou Dodo^{2,4}, Poka-Mayap Virginie^{2,3}

¹ Département de Médecine Interne et Spécialités, Faculté de Médecine et des Sciences Biomédicales, Université de Yaoundé 1, Yaoundé, Cameroun

² Association Pour la Recherche, l'Éducation et la Santé (APRES), Yaoundé, Cameroun

³ Hôpital Jamot de Yaoundé, Yaoundé, Cameroun

⁴ Faculté de Médecine et des Sciences Biomédicales, Université de Ngaoundéré, Garoua, Cameroun

TO CITE: Pefura-Yone EW et al. **Description des données en statistique et biostatistique : Types de variables.** *The Papers of Medical Sciences* 2021;2:e001.

KEYS WORDS:

data, variables, statistics, biostatistics

ARTICLE INFO

Received: 11th January 2021

Accepted: 22th April 2021

Available online: 27th April 2021

***Correspondence to:**

Pefura-Yone EW, Email: pefura2002@yahoo.fr

ISSN: 2663-7545

Copyright ©2021, Pefura-Yone et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution and reproduction in any medium, provided the original authors and source are credited.

RESUME

L'analyse des données en statistique s'appuie sur le type de données ou de variables à analyser. Les variables qualitatives et quantitatives ne sont pas résumées et analysées de la même façon. La classification des variables à quatre niveaux est la classification la plus usitée et permet de séparer les variables en fonction de leur type, de leur échelle et de leur nature.

Mots-clés : données, variables, statistique, biostatistique

ABSTRACT

Data analysis in statistics is based on the type of data or variables to be analyzed. Qualitative and quantitative variables are not summarized and analyzed in the same way. The classification of variables at four levels is the most commonly used classification and allows variables to be separated according to their type, scale and nature.

Key-words: data, variables, statistics, biostatistics

INTRODUCTION

L'analyse des données en statistique et biostatistique est primitivement basée sur la catégorisation des variables. Une variable en statistique est une caractéristique susceptible de prendre une valeur différente selon les individus composant la population d'étude ou l'échantillon [1]. Les variables sont parfois appelées données. Chaque individu composant la population d'étude est appelé unité statistique. Les méthodes statistiques employées dépendent étroitement des types de variables à analyser. Par exemple, le genre est la caractéristique permettant de classer les individus en homme ou femme. On ne peut pas calculer la valeur moyenne du genre puisque l'interprétation du résultat n'aura aucun sens logique. De l'autre côté, dans une population humaine de tout âge, il serait fastidieux de donner la fréquence de chaque âge ; ceci ne sera pas facile à interpréter. Ainsi, la variable genre (hommes ou femmes) ne peut pas répondre à une mesure et ses modalités ne peuvent être que qualifiées : il s'agit d'une variable qualitative. Par contre l'âge est dénombrable (1an, 2, 3...120 ans) ou alors l'âge de chaque individu est mesurable : il s'agit d'une variable quantitative. Nous allons décrire les variables en précisant leurs caractéristiques générales, les types de variables, les échelles de mesure des variables et la nature des données.

1. CARACTÉRISATION D'UNE VARIABLE

Les variables sont caractérisées par [2] :

- Leur type (qualitatif ou quantitatif) ;
- Leur échelle (nominale, ordinale, intervalle, rapport) ;
- Leur nature (discrète, continue).

2. TYPES DE VARIABLES

Il existe deux types de variables. Les variables qualitatives et les variables quantitatives [1–4].

Une variable qualitative est une variable dont les modalités ne sont pas dénombrables mais pouvant être classées. Les valeurs mesurées sur chaque modalité ne sont pas quantifiables. Les modalités d'une variable qualitative sont simplement qualifiables et peuvent être rangées dans plusieurs catégories. Les variables qualitatives sont ainsi aussi appelées variables catégorielles. Voici quelques exemples de variables qualitatives : sexe (homme, femme), groupe ethnique (bantou, soudano-sahélien, mixte), niveau d'éducation (non scolarisé, primaire, secondaire, supérieur). Quand la variable contient deux modalités elle est dite **dichotomique** ou **binomiale** (variable sexe), sinon elle est dite **polytomique** (variable niveau d'éducation). Une variable dichotomique constituée uniquement des modalités 0 et 1 est une variable **binaire**.

Une variable quantitative est une variable sur laquelle il est possible de réaliser les calculs numériques et ainsi faire des mesures ou quantifier les valeurs. Par exemple, la taille, les paramètres spirométriques ou de la fonction respiratoire (volume expiratoire maximal à la 1ère seconde, VEMS ; capacité vitale forcée, CVF, rapport VEMS/CVF, débit expiratoire moyen médian, DEMM) et l'âge des individus sont des variables quantitatives. L'âge en années peut être transformé en âge en mois (en divisant l'âge en années par 12) : on obtiendra une autre quantité pouvant aussi se prêter aux calculs numériques.

3. ECHELLES DE MESURE DES VARIABLES

Les échelles de mesure permettent de mieux caractériser les types de variables [2, 3, 5]. Les variables qualitatives peuvent prendre deux formes d'échelles de mesure : variable qualitative **nominale** ou variable qualitative **ordinaire**. Une variable nominale est caractérisée par les modalités ou catégories non hiérarchiques ou non classables [sexe (les hommes ne sont pas supérieurs aux femmes et vice-versa, groupe ethnique (aucun ordre de classement n'est possible)]. Une variable ordinaire répond à des modalités obéissant à un classement ordonné ou rangé allant du plus bas au plus élevé ou inversement (niveau d'éducation : non scolarisé, primaire, secondaire, supérieur). Les variables répondant à la classique **échelle de Likert** (pas du tout d'accord=1 ... tout à fait d'accord=4) sont des variables ordinaires [3].

Les échelles de mesure applicables aux données quantitatives sont également de deux ordres : échelle d'intervalle et échelle de rapport. Pour **l'échelle d'intervalle**, les écarts sont objectifs et la valeur zéro est une valeur arbitraire ne signifiant pas l'absence de valeur. La différence entre deux valeurs aura un sens mais le calcul du rapport entre les valeurs n'a aucun sens logique. Par exemple la température d'évaluation des paramètres spirométriques exprimée en °C est une variable quantitative d'intervalle: une température de 37°C est plus élevée qu'une température de 36°C mais l'écart entre 37°C et 36°C est le même qu'entre 36°C et 35°C et 0°C ne signifie pas l'absence de température mais bel et bien une température de 0°C). **L'échelle de rapport** répond à des écarts proportionnels entre les mesures mais le zéro est universel ou est un vrai zéro. Le calcul du rapport entre deux valeurs a un sens et autorise une interprétation logique. Un exemple de variable de rapport est la variable de comptage du nombre d'habitant dans les zones de dénombrement (0 habitant signifie pas d'habitants, la population de la zone A est le double de celle de la Zone B a un sens). Parfois la différence entre les types et les échelles de variables ne sont pas évidentes et certaines variables paraissent ambiguës. Dans cette situation, il faut comprendre dans quelles circonstances ces variables seront appliquées.

La caractérisation d'une variable par son type et son échelle est dénommée en statistique classification des données à 4 niveaux ou de **Stevens [5]** (Figure 1).

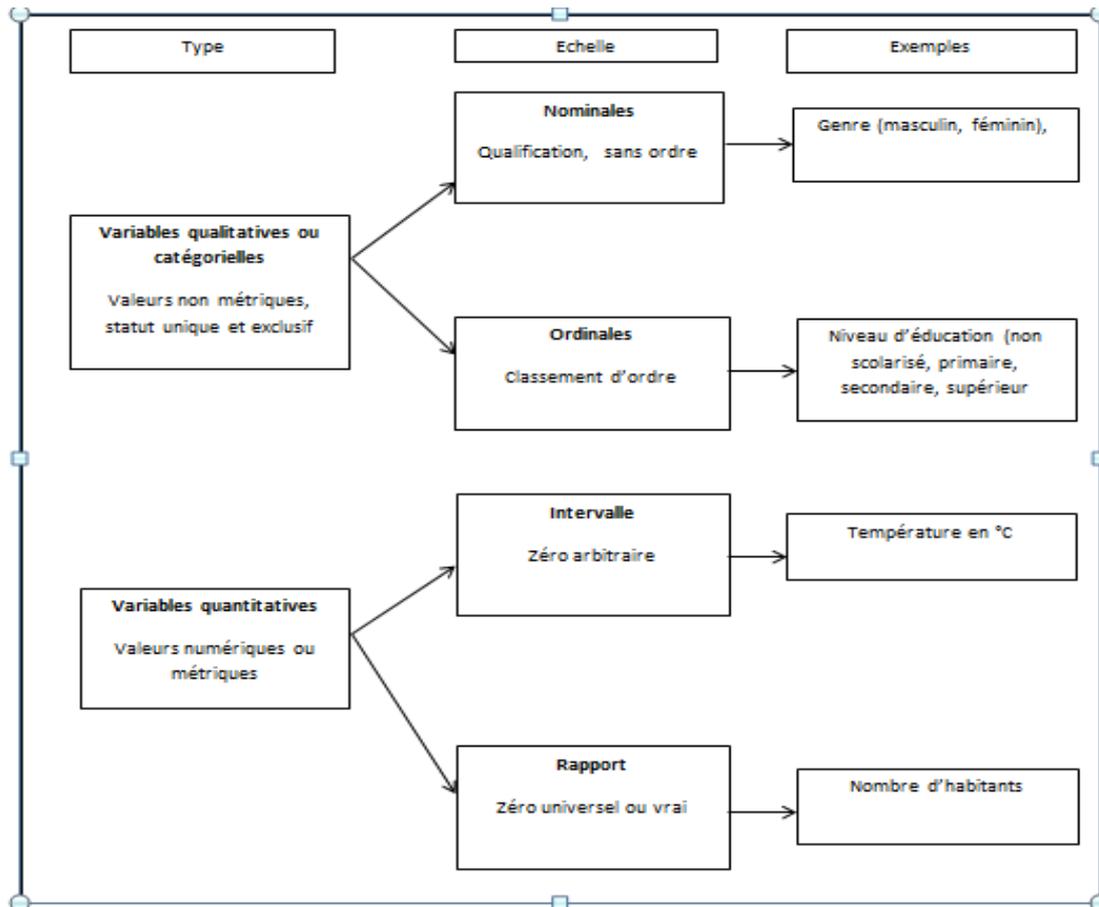


Figure 1 : Classification à 4 niveaux des variables

4. NATURE DES DONNÉES

La nature des données s'applique aux variables quantitatives [3]. Les variables quantitatives peuvent être discrètes ou continues. **Une variable quantitative discrète** prend des valeurs finies dans un intervalle de valeurs (par exemple : Nombre d'habitants des zones de dénombrement, 100, 101, 102, 103, 104, 105 ; entre 100 et 105 nous avons 6 valeurs numériques possibles). **Une variable quantitative continue** peut prendre toutes les valeurs possibles (infinies) entre deux valeurs sans discontinuer (par exemple infinité de taille possible entre 150 et 155 cm).

CONCLUSION

La compréhension de la classification des variables est l'étape essentielle initiale à l'analyse des données. La classification des données à 4 niveaux permet de choisir le type d'analyse statistique applicable à chaque variable. La description des données à l'aide des outils de la statistique descriptive sera abordée dans le prochain article.

REFERENCES

1. Ancelle T. Statistique Epidémiologie. 3è édition. 2015. 308 pages.

2. Motulsky H. Types of variables. In: *Intuitive biostatistics: A Nonmathematical Guide to Statistical Thinking*, Second edition. 2010. p. 67–70.
3. Broc G. Panorama sur les données. In: *Stats faciles avec R. Guide pratique*. 2016. p. 19–29.
4. Velleman PF, Wilkinson L. Nominal, Ordinal, Interval, and Ratio Typologies Are Misleading. *Am Stat*. 1993;47:65. doi:10.2307/2684788.
5. Stevens SS. On the Theory of Scales of Measurement. *Science* (80-). 1946;103:677–80. doi:10.1126/science.103.2684.677.