

# Statistiques descriptives

Les statistiques descriptives se contentent, comme leur nom l'indique, de décrire les variables étudiées dans une population. Elles permettent ainsi d'obtenir des informations précieuses sur la distribution statistique de chaque variable étudiée dans une recherche scientifique.

Il existe 2 grandes approches permettant de décrire les variables d'une population étudiée : l'approche graphique (histogramme, diagramme de fréquences, diagramme cumulatif, etc.) et l'approche numérique (moyenne et écart-type, médiane, percentiles, fréquences, etc.).

## Approche numérique :

### 1. Moyenne et écart-type

La **moyenne** (*mean*), notée  $\bar{x}$ , est la valeur unique que devraient avoir tous les individus d'une population (ou d'un échantillon) pour que leur total soit inchangé (*Wikipédia, 2018*). La moyenne arithmétique est donc donnée par la formule suivante :

$$\frac{\text{somme de toutes les valeurs prises par la variable étudiée}}{\text{nombre d'individus dans l'échantillon étudié}}$$

La moyenne arithmétique est en fait un paramètre de *position* car elle se situe exactement au centre de la distribution statistique des données de la variable étudiée. Il faut savoir également que la moyenne arithmétique est particulièrement sensible aux valeurs aberrantes car elle prend en compte toutes les valeurs de l'échantillon. Une donnée particulièrement haute (ou basse) peut donc fausser la moyenne et son interprétation.

L'**écart-type** (*standard deviation = SD*) est toujours présenté avec la moyenne. Il fournit des informations sur la variabilité des données de l'échantillon autour de la moyenne. Il s'agit donc d'un indicateur de *dispersion*. Il donne en fait des indications sur la variabilité des données récoltées pour une variable.

Dans les publications scientifiques, les résultats de moyenne et d'écart-type sont présentés comme suit : moyenne  $\pm$  écart-type ou, la plupart du temps, ***mean  $\pm$  SD***.

La moyenne s'applique aux variables quantitatives.

### 2. Médiane et percentiles

La **médiane** (*median*), notée  $M$ , divise une distribution d'observations en 2 parties égales. Elle constitue donc la valeur centrale par excellence. Elle est la valeur qui laisse 50% des observations en dessous d'elle et 50% des observations au-dessus d'elle.

Contrairement à la moyenne, la médiane est particulièrement robuste car elle ne tient pas compte des valeurs extrêmes de la distribution des données.

Sont associés à la médiane les **percentiles**, nommés *P*. Par définition, un percentile est la valeur *X* de l'échantillon qui laisse *X*% des observations en-dessous et (100-*X*%) des observations au-dessus d'elle. Par exemple, le percentile 30 (*P*30) est la valeur de l'échantillon qui laisse 30% des observations en-dessous et 70% des observations au-dessus d'elle.

La médiane correspond donc au percentile 50 (**P50**). Elle est présentée, dans les articles scientifiques, avec les valeurs des percentiles 25 et 75 : **Median (P25 – P75)** ou *M* (*P*25 – *P*75).

La médiane s'applique aux variables quantitatives.

### 3. Fréquence

La **fréquence** correspond au nombre de fois qu'une donnée est observée ou encore au nombre de fois qu'un phénomène est observé. La fréquence peut s'exprimer selon 2 façons :

- La **fréquence absolue** : elle s'exprime en chiffres et correspond simplement à l'effectif des observations d'une variable étudiée ;
- La **fréquence relative** : elle s'exprime en pourcentage et correspond au rapport  $\frac{\text{nombre d'effectifs de l'observation étudiée}}{\text{nombre total d'observations récoltées}}$ .

Dans les publications scientifiques, les deux formes se rencontrent. Il est fréquent de voir l'appellation suivante : ***n* (%)** correspondant d'abord au nombre « brut » d'observations et, ensuite, à son pourcentage par rapport à la population totale étudiée.

La fréquence s'applique aux variables qualitatives, aux variables quantitatives discrètes ou encore aux variables binaires.

### 4. Exemples de présentation

Table 1: Demographic profile of the survey respondents (N=349).

Variable	Mean ± SD	n (%)
Age (years)	31.6 ± 4.65	
Sex		
Men		158 (45.3)
Women		191 (54.7)
Marital status		
Married		147 (42.1)
In couple or in legal cohabitation		191 (54.7)
Single		3 (0.86)
Divorced		2 (0.57)
Separated		5 (1.43)
Other		1 (0.29)

*Source* : Albert A. Biostatistique. Les Éditions de la Province de Liège et les Editions de l'Université de Liège, Liège, 2005.