



THÉMATIQUE 1 - ANALYSE UNIVARIÉE

Data Sciences Revision

MENTION MATHÉMATIQUES ET INFORMATIQUE
PARCOURS HPDA

10/09/2024

Rédigé par :

PAULY ALEXANDRE

alexandre.pauly@cy-tech.fr

1 Différencier les variables

- *Variable qualitative* : La variable est dite qualitative quand les modalités sont des catégories.
 - *Nominale* : La variable est dite qualitative nominale quand les modalités ne peuvent pas être ordonnées (*ex: le sexe, le genre, etc.*).
 - *Ordinale* : La variable est dite qualitative ordinale quand les modalités peuvent être ordonnées. Le fait de pouvoir ou non ordonner les modalités est parfois discutable (*ex: catégories socioprofessionnelles, on admet d'ordonner les modalités : 'ouvriers', 'employés', 'cadres'. Si on ajoute les modalités 'sans profession', 'enseignant', 'artisan', l'ordre devient beaucoup plus discutable.*)
- *Variable quantitative* : Une variable est dite quantitative si toutes ses valeurs possibles sont numériques.
 - *Discrète* : Une variable est dite discrète, si l'ensemble des valeurs possibles est compris dans un intervalle fermé (*ex: les jours d'un mois, note, etc.*).
 - *Continue* : Une variable est dite continue, si l'ensemble des valeurs possibles est continu (*ex: la taille, le poids, etc.*).

Attention : Certains cas sont à relativiser, l'âge est théoriquement une variable quantitative continue, mais en pratique, l'âge est mesuré dans le meilleur cas au jour près.

2 Représenter les données

2.1 Variable qualitative nominale

Un diagramme en secteurs permet de représenter les fréquences, alors qu'un diagramme en barres permet de représenter les effectifs.

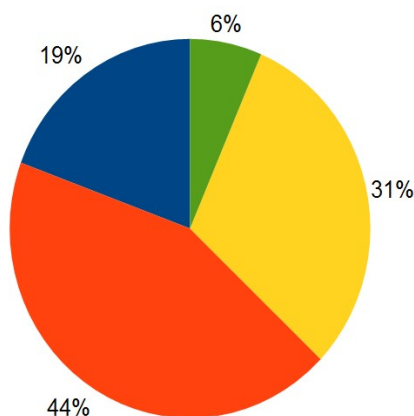


Figure 2: Diagramme en secteurs

Attention : Penser à mettre une légende lorsque nécessaire et **OBLIGATOIREMENT** un titre

2.2 Variable qualitative ordinale

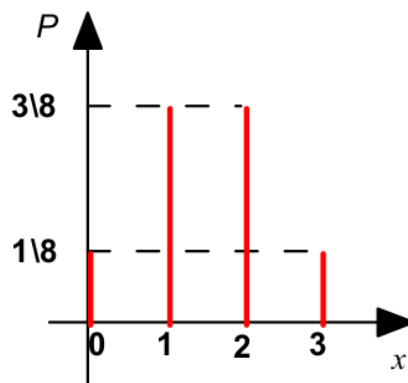
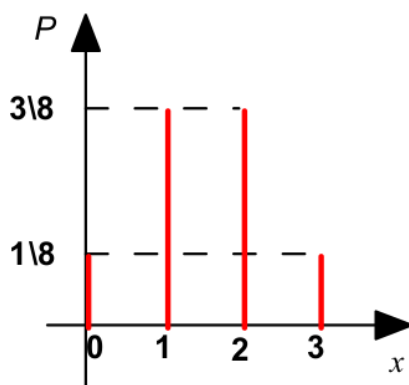


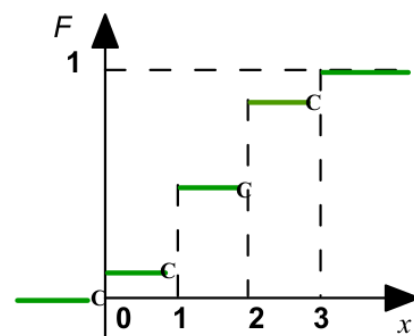
Figure 1: Diagramme en batons

Attention : Penser à mettre une légende lorsque nécessaire et **OBLIGATOIREMENT** un titre

2.3 Variable quantitative discrète



(a) Diagramme en baton des effectifs

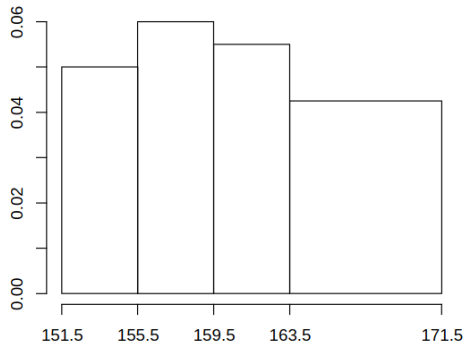


(b) Fonction de répartition

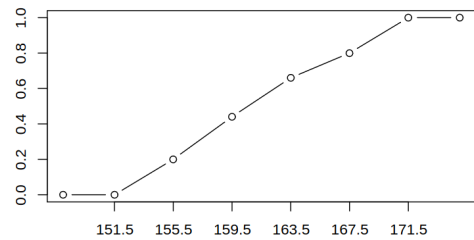
Figure 2: Représentation de variables quantitative discrète

Attention : Penser à mettre une légende lorsque nécessaire et **OBLIGATOIREMENT** un titre

2.4 Variable quantitative continue



(a) Histogramme des fréquences



(b) Fonction de répartition d'une distribution groupée

Figure 3: Représentation de variables quantitative continue

2.5 Boîte de Tukey

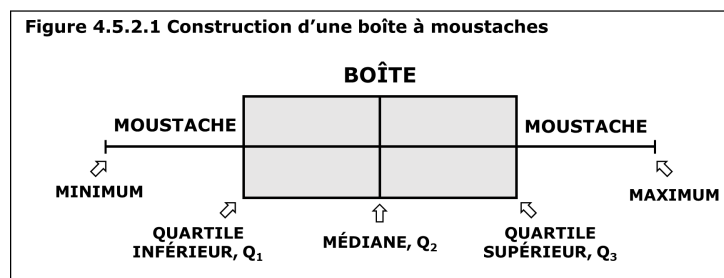


Figure 4: Boîte de Tukey

Un diagramme en barres permet de représenter les effectifs.

Attention : Penser à ajouter la moyenne et autres indicateurs par soucis de facilité de lecture.

3 Indicateurs de position et de dispersion

Les quatre indicateurs de position : le mode, la médiane, le quartile et la moyenne
Comparaison de la médiane et de la moyenne d'une même série.

Les quatre indicateurs de dispersion : l'étendue, l'intervalle interquartile, la variance et l'écart type.

4 Variables centrées-réduites

L'objectif est de pouvoir mieux comparer les variations. Cela équivaut à un changement d'unité, et n'a pas d'incidence sur les profils de variation.

Rappel : Les caractéristiques d'une variable centrée-réduite sont une espérance nulle ainsi qu'un écart type de 1.