

85 → Programmation en Python

But → AI (Data Analyst ?)

2 CC : • Chapitre 1, 2 & 3
• Chapitre 4 & 5

→ Exercices type TD
→ Screen de TP à expliquer

Formulaire = QUE DES FORMULES
+ Calculatrice autorisées

STATISTIQUES → quantitatif & qualitatif
Numerical & Categorical
(valeurs → nombre fini ou infini mais dénombrable de valeurs) (nominal → couleurs, nombre fini de valeurs ordinal → ordonnées, classes, ...)

→ discret = fini ou infini dénombrable ∑: Somme
continu = valeurs infinies ?
↳ Numerical (Les valeurs possibles sont des nombres) ∫: Intégrale

Unidimensionnelle → 1 variable (types de données ? localisation (→ tendance centrale) ? dispersion ?)

Bidimensionnelle → 2 variables

→ localisation :

Pour les variables continues :

âge car nombre réel (20, 35 ans)

Plusieurs moyens de décrire la donnée → graphique
→ numérique

• Moyenne arithmétique : $\bar{x} = \frac{1}{n} \sum x$ résumé
△ ! préciser
→ toutes les valeurs de x

Il existe la moyenne géométrique, harmonique, ...

Tendance centrale

• Médiane : $Q_2 =$ si n impair, $x_k = x_{(\frac{n+1}{2})}$
↳ Statistique d'ordre (valeurs ordonnées)
 $x_{(k)} =$ k^{ème} valeur la + petite

si n pair, $x_k = [x_{(n/2)}, x_{(n/2+1)}]$ bornes exclues

$\frac{x_{(n/2)} + x_{(n/2+1)}}{2}$

△ logiciel = interpolation donc résultat différents

Variables qui ont la loi qui sont discrètes, pas la loi

Valeur la + grande et la + petite ≠ min, max

• Valeur la + grande et la + petite

• Valeur la + observée → Mode : $x^* = \underset{x}{\operatorname{argmax}} n(x)$
 x^* tel que $(t_j) f(x^*) = \max_x f(x)$
→ valeur la + observée

Pour les variables discrètes :

• Mode (ordinal & nominal)

• Médiane (ordinal seulement)

• Moyenne (ordinal seulement mais pas utilisée)

→ dispersion :

Mesure de la variabilité des valeurs observées (indispensable à l'IA)

Pour les variables continues :

• Ecart-Type : $\sigma^2 = \frac{1}{n} \sum (x - \bar{x})^2$

Standard Deviation

→ Variance = σ^2

⚠ Calculer σ^2 puis la racine

$$= \frac{1}{n} \sum x^2 - \bar{x}^2$$

Moyenne des carrés moins carré de la

moyenne

moins de complexité

• Etendue : $R = |x_{(n)} - x_{(1)}|$

Range

• Quartile : $Q_{np} = x_{(np)}$

pourcentile parfois

$$Q_{np} = [x_{(Lnp-1)}, x_{(Lnp+1)}]$$

$x_{(np)}$ est entier ou pas

x

$\lceil \rceil$
 $\lfloor \rfloor$
 $\lfloor \rfloor$

partie entière

partie entière supérieure

partie entière inférieure

• Coefficient de variation : $CV = \frac{\sigma}{\bar{x}}$ sans unité

$\alpha = \frac{1}{2}$
par défaut

• Etendue inter-quartile : $IQR = |Q_3 - Q_1|$

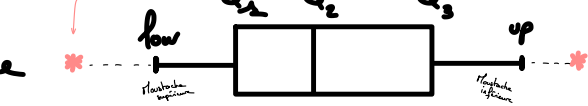
50 % de valeurs centrales

• Epurant : $R_\alpha = [Q_1 - \alpha_{IQR}, Q_3 + \alpha_{IQR}]$

Range d'ordre alpha

valeur observée immédiatement plus petite que $Q_1 \rightarrow -$ + grande que Q_3

outlier
valeur + petite que low



Box plot = boîte à moustache