

TP 1 : statistique descriptive à une dimension

0.1 Rappel théorique

0.1.1 Définitions

Tout le long du rappel théorique, nous illustrerons les définitions grâce à un exemple.

n données issues d'une variable X sont observées.

exemple : nous observons la taille de 5 personnes, n vaut donc 5, et les tailles respectives observées sont 182 cm, 175 cm, 182 cm, 185 cm et 190 cm.

p données différentes (!) sont observées, qui, classées dans l'ordre croissant, sont $x_1, x_2, \dots, x_p$.

exemple : $x_1 = 175$ cm, $x_2 = 182$ cm, $x_3 = 185$ cm, $x_4 = 190$ cm.

Si le nombre total d'observations, n , est très grand, nous regrouperons nos observations en classes.

exemple : $C_1 = [175 \text{ cm}, 180 \text{ cm}]$, $C_2 =]180 \text{ cm}, 185 \text{ cm}]$, $C_3 =]185 \text{ cm}, 190 \text{ cm}]$.

n_i = fréquence absolue associée à x_i (respectivement C_i)
= nombre d'observations égales à x_i (respectivement, appartenant à C_i).

exemple :

- pour les observations x_i : $n_1 =$ fréquence absolue de $x_1 = 1$; $n_2 = 2$;
- pour les classes C_i : $n_1 =$ fréquence absolue de $C_1 = 1$; $n_2 = 3$.

$N_i = \sum_{j=1}^i n_j$ = fréquence absolue cumulée associée à x_i (respectivement C_i).
= nombre d'observations plus petites ou égales à x_i (respectivement appartenant à C_i et aux classes précédentes).

exemple :

- pour les observations x_i : $N_1 =$ fréquence absolue cumulée de $x_1 = 1$; $N_2 = 3$;
- pour les classes C_i : $N_1 =$ fréquence absolue cumulée de $C_1 = 1$; $N_2 = 4$.

$f_i = \frac{n_i}{n}$ = fréquence relative associée à x_i (respectivement à C_i).

exemple :

- pour les observations x_i : f_1 = fréquence relative de $x_1 = 1/5$; $f_2 = 2/5$;
 - pour les classes C_i : N_1 = fréquence relative de $C_1 = 1/5$; $f_2 = 3/5$.
- $F_i = \sum_{j=1}^i f_j = \frac{N_i}{n}$ = fréquence relative cumulée associée à x_i (respectivement à C_i).

exemple :

- pour les observations x_i : F_1 = fréquence relative cumulée de $x_1 = 1/5$; $F_2 = 3/5$;
- pour les classes C_i : F_1 = fréquence relative de $C_1 = 1/5$; $F_2 = 4/5$.

Continuons à illustrer les différents paramètres uniquement grâce aux données x_i (pas les classes).

0.1.2 Paramètres de position

1. $\bar{x}$ = moyenne arithmétique = $\frac{1}{n} \sum_{i=1}^p n_i x_i$

exemple : $\bar{x} = \frac{1}{5}(175 + 2 \cdot 182 + 1 \cdot 185 + 1 \cdot 190)$

2. $\tilde{x}$ = médiane = nombre qui laisse la moitié des observations à sa gauche ainsi qu'à sa droite (aussi appelé deuxième quartile).

exemple : $\tilde{x} = x_3 = 182$

0.2 Paramètres de dispersion

1. s^2 = variance = $\frac{1}{n} \sum_{i=1}^p n_i (x_i - \bar{x})^2$.

2. écart-type = $s = \sqrt{s^2}$.

3. premier quartile = nombre qui laisse 1/4 des observations à sa gauche.

exemple : premier quartile = $x_2 = 182$ cm

4. troisième quartile = nombre qui laisse 3/4 des observations à sa gauche.

exemple : troisième quartile = $x_3 = 185$ cm

5. intervalle interquartile = [premier quartile, troisième quartile].

exemple : [182 cm , 185 cm]

0.2.1 Paramètres de dissymétrie

le paramètre de dissymétrie est $m_3 = \frac{1}{n} \sum_{i=1}^p n_i (x_i - \bar{x})^3$. Grâce à cela, nous construisons le paramètre de Fisher, noté g_1 et défini par $g_1 = m_3/s^3$.

0.2.2 Paramètres d'aplatissement

Le paramètre d'aplatissement est $m_4 = \frac{1}{n} \sum_{i=1}^p n_i (x_i - \bar{x})^4$. Grâce à cela, nous construisons le paramètre de ..., noté g_2 et défini par $g_2 = (m_4/s^4) - 3$.