MÉDIANE (C5, F3)
(19 / 12 / 2019, © Monfort, Dicostat2005, 2005-2019)

La **médiane** est un quantile particulier qui joue un rôle important en **Statistique** en tant que « valeur représentative » d'une distribution, d'une **variable aléatoire**, ou encore d'un ensemble d'observations : elle joue ainsi un rôle de **caractéristique** ou **paramètre** de **centralité**.

(i) Soit $(\Omega, \mathcal{T}, P)$ un **espace probabilisé** et $\xi : \Omega \mapsto \mathbf{R}$ une **vars**.

On appelle **valeur médiane**, ou **valeur équiprobable**, ou simplement **médiane, (théorique)** de ξ (ou de sa **loi** P^{ξ}) tout nombre réel, noté $Q_{1/2}\,\xi$ ou encore M_{ξ} , tq :

(1) $\quad P(\xi \leq Q_{1/2}\,\xi) \geq 1/2 \quad$ et $\quad P(\xi \geq Q_{1/2}\,\xi) \leq 1/2$,

ie tout **quantile** d'ordre 1 / 2 de P^{ξ}.

L'ensemble $\mathcal{E}$ des valeurs $Q_{1/2}\,\xi \in \mathbf{R}$ vérifiant (1), ie l'ensemble des valeurs médianes, est appelé **intervalle médian**, **partie médiane** ou encore **ensemble médian**, de P^{ξ} (ou de ξ, ou aussi de R).

Si F est la **fr** associée à P^{ξ}, la définition (1) s'écrit encore :

(2) $\quad F(Q_{1/2}\,\xi) \geq 1/2 \quad$ et $\quad F(Q_{1/2}\,\xi) \leq 1/2$.

En particulier, si F est continue, $\mathcal{E}$ se réduit à un seul point : $Q_{1/2}\,\xi$ est alors unique. Dans ce cas, (1) ou (2) devient :

(3) $\quad F(Q_{1/2}\,\xi) = 1/2$.

(ii) On montre que :

(a) $\forall \lambda \in \mathbf{R}$, on a $Q_{1/2}(\lambda\,\xi) = \lambda \,.\, Q_{1/2}\,\xi$;

(b) l'**écart absolu moyen** (théorique) $E\,|\xi - \alpha|$ est minimum pr à α lorsque $\alpha = Q_{1/2}\,\xi$ (propriété comparable à celle de l'**espérance** pr à la **variance**) ;

(c) si $\xi \in \mathcal{L}_{\mathbf{R}}^{2}(\Omega, \mathcal{T}, P)$ est de carré intégrable, alors l'**écart** entre la médiane et l'espérance est borné par l'**écart-type** :

(4) $\quad |Q_{1/2}\,\xi - E\,\xi| \leq (2 \,.\, V\,\xi)^{1/2}$.

(iii) Soit $X = (X_1, \ldots, X_N) : \Omega \mapsto \mathbf{R}^N$ un **échantillon iid** selon P^{ξ}.

On appelle **valeur médiane**, ou **valeur équiprobable**, ou simplement **médiane**, **(empirique)** de ξ (ou de X) la médiane $q_{1/2}\,X$ de la **loi empirique** P_N associée à X, ie aussi le **quantile empirique** d'ordre p = 1 / 2 de X. On a donc :

$$\in [X^{(N/2)}, X^{((N/2)+1)}] \qquad \text{si } N \in 2\,\mathbf{N},$$

(5) $q_{1/2}$ X ou m_N X

$= X^{((N/2)+1)}$ si $N \in 2\mathbf{N} + 1$,

où $X^{(.)} = (X^{(1)}, ..., X^{(N)})$ est l'échantillon ordonné (cf **statistique ordonnée**) associé à X.

La définition de la médiane empirique est donc analogue à celle de la médiane théorique : on remplace P^{ξ} (resp F) par la loi (resp fr) empirique P_N (resp F_N) associée à X dans (1) (resp (2)) (cf **statistique naturelle**).

La médiane n'est donc pas, en général, une **caractéristique** unique. Lorsque $N \in 2\mathbf{N}$, on remplace souvent, en pratique, $q_{1/2}$ X par la **moyenne** simple suivante (cf **interpolation**) :

(6) $q_{1/2}'$ X ou m_N' X $= (X^{(N/2)} + X^{((N/2)+1)}) / 2$,

ce qui permet de la définir de façon unique.

(iv) On peut aussi définir la **médiane empirique** $q_{1/2}$ X comme la « valeur » rendant minimum (pr à m) la fonction :

(7) $\varphi(m) = d_1(X, m\, e_N) = \Sigma_{n=1}^{N} |X_n - m|$,

ie la **distance** associée à une **norme** usuelle de $\mathbf{R}^N$.

Ainsi, $q_{1/2}$ X existe toujours, mais n'est pas nécessairement unique.

(v) On montre que la médiane (théorique) de la médiane empirique $q_{1/2}$ X (considérée comme **vars** ou **statistique**) est la médiane (supposée unique) de la **loi** P^{ξ}, ie :

(8) $Q_{1/2}(q_{1/2} X) = Q_{1/2}\, \xi$.

(vi) La médiane empirique constitue un **estimateur** « naturel » de la médiane théorique.

Comme elle ne dépend pas des **valeurs extrêmes** de X dès que sa taille dépasse 2 (N > 2), elle joue un rôle particulier (cf aussi **forme légale**) :

(a) en théorie de la **robustesse** : eg estimation robuste d'une **caractéristique** de **centralité**, ou d'un **paramètre de position** ;

(b) en théorie de la sensibilité : cf **courbe de sensibilité**, **courbe d'influence**.

(vii) La notion de médiane, et notamment la médiane empirique, s'associe directement à celle de **régression quantilaire**, qui intervient dans l'étude du **modèle de régression** (**régression robuste**).

(viii) On peut étendre la définition d'une médiane, eg :

(a) à une **va** $\xi : \Omega \mapsto E$ à valeurs dans un **ensemble** ordonné $(E, \leq)$, dès lors que la **lp** ou la **fr** sont définies (cf **relation d'ordre**).

(b) à un **vecteur aléatoire** $\xi : \Omega \mapsto \mathbf{R}^K$ (cf **médiane multidimensionnelle**).