Société de Calcul Mathématique SA
Outils d'aide à la décision
depuis 1995

Utilisations de la médiane et de la "Median Absolute Deviation"

Bernard Beauzamy

mars 2019

1. La médiane

Soit X une variable aléatoire, dont on a observé un échantillon x_n, $n = 1, ..., N$.

La médiane désigne le nombre m tel qu'il y ait autant de valeurs de l'échantillon au-dessus qu'au-dessous. Cela n'a de sens que pour une v.a. réelle (et cela n'en a pas pour une variable complexe, ou qualitative, par exemple). Il y a un problème de définition si le nombre de valeurs N est pair : on se retrouve avec un intervalle (en général, on prend le milieu de cet intervalle).

Une fois la valeur de m déterminée, elle est invariante par éloignement des valeurs de l'échantillon : les valeurs au-dessus de m peuvent s'éloigner à droite, les valeurs au-dessous peuvent s'éloigner à gauche, sans que cela change la médiane.

On peut dire grossièrement que la médiane est un concept qualitatif, permettant de partager les valeurs de l'échantillon en deux classes d'égale importance ; ce n'est pas un concept quantitatif. Elle ne doit pas être utilisée s'il s'agit de proposer une réglementation ou une norme (fixer un seuil, par exemple).

2. Median Absolute Deviation

Elle est définie par : $mad = med|X - m|$

Pour comprendre ce que cette formule représente, commençons par le cas simple où $m = 0$: X a autant de valeurs positives que négatives.

On regarde $|X|$; les valeurs sont d'une part les $x_i > 0$ et d'autre part les $y_i = -x_i$ pour les $x_i < 0$.
La *mad* sera le nombre A tel qu'il y ait autant de x_i, y_i de part et d'autre.

Mais :

– Être au-dessus de A pour $|X|$ signifie pour X être $< -A$ ou $> A$;

– Être entre 0 et A pour $|X|$ signifie pour X être dans l'intervalle $-A, A$.

Par conséquent, la *mad* est le nombre A tel que (pour X) il y ait autant de valeurs dans l'ensemble $]-\infty, -A] \cup [A, +\infty[$ que dans l'intervalle $[-A, A]$.

Si on revient au cas général, il faut appliquer ceci à $X - m$.

Concrètement, pour déterminer la *mad*, on part de $X - m$ et on regarde $X - m \pm \alpha$; on commence avec $\alpha = 0$ et on augmente $\alpha > 0$ progressivement. On s'arrête quand le nombre de points dans l'intervalle $[X - m - \alpha, X - m + \alpha]$ est égal au nombre de points hors de cet intervalle.

Nous constatons que ce n'est pas une vraie mesure de dispersion. On peut éloigner les points les plus grands ou les plus petits sans changer le résultat. Plus généralement, la *mad* est invariante par modification des valeurs à l'intérieur des intervalles. Les points du premier intervalle peuvent être tous mis à m et les points du second ensemble éloignés indéfiniment (et même remplacés par leurs opposés) sans que cela change A. Là encore, c'est un concept qualitatif et non quantitatif, qu'il ne faut pas utiliser comme base réglementaire.