Université de Poitiers
Département de Mathématiques

Statistique descriptive, 1er semestre, année univ. 2009-2010 **Fiche 2**

Statistiques descriptives univariées, suite

Exercice 1

Soit un échantillon $x = (x_1; x_2; \ldots; x_n)$ pondéré par les poids $p = (p_1; \ldots; p_n)$. On suppose que les p_i sont strictement positifs et de somme 1.

1) Montrer que la fonction
$$J_2(c) := \sum_{i=1}^{n} p_i (x_i - c)^2$$
est minimale pour $c = \bar{x}$ la moyenne arithmétique empirique de x et que le minimum vaut $s^2(x)$ la variance empirique de l'échantillon x.

2) On munit $\mathbb{R}^n$ du produit scalaire $< x; y >= \sum_{i=1}^{n} p_i x_i y_i$. Montrer que la projection de x sur la droite engendrée par le vecteur $\underline{1} = (1; \ldots; 1)$ est $\bar{x}$ puis que la distance entre x et cette droite est $s(x)$.

3) Montrer que $s^2(x) = 0$ si et seulement si tous les x_i sont égaux. Dans ce cas, que vaut $\bar{x}$?

4) Montrer les propriétés suivantes :
$$\overline{ax + b\underline{1}} = a\bar{x} + b, \quad s^2(ax + b\underline{1}) = a^2 s^2(x).$$
Si $s^2(x) \neq 0$, que valent la moyenne (arithmétique) et la variance de l'échantillon centré réduit $x' = \frac{x - \bar{x}}{s(x)}$?

Exercice 2

Soit un échantillon $x = (x_1; \ldots; x_n)$ pondéré par les poids $p = (p_1; \ldots; p_n)$. On suppose que les p_i sont strictement positifs et de somme 1.

1) Quels sont les points qui minimisent la fonction
$$J_1(c) := \sum_{i=1}^{n} p_i |x_i - c| \ ?$$
On pourra introduire la statistique d'ordre $x_{(\cdot)}$ de x.

2) Que vaut la médiane $\widetilde{ax + b\underline{1}}$ de l'échantillon $ax + b\underline{1}$?

Exercice 3

Soit un échantillon $x = (x_1; \ldots; x_n)$ pondéré par les poids $p = (p_1; \ldots; p_n)$. Pour $y \in \mathbb{R}$ on note $x(y) = (x_1; \ldots; x_{n-1}; y)$ l'échantillon obtenu à partir de x en remplaçant la dernière observation x_n par y. On interprète $x(y)$ comme étant l'échantillon x au sein duquel une erreur de mesure s'est glissée.

1) Étudier la fonction $y \mapsto \overline{x(y)}$. On donnera notamment sa dérivée et les limites en $\pm\infty$.

2) Étudier de même la fonction $x \mapsto \widetilde{x(y)}$ (médiane de $x(y)$). On pourra pour simplifier supposer qu'une somme de p_i ne vaut jamais $\frac{1}{2}$ si bien que la médiane est définie sans ambiguité.

3) Commenter : lequel de ces deux indicateurs de position vous semble le moins sensible aux valeurs aberrantes ?

4) Pensez-vous que l'écart type d'un échantillon soit sensible aux valeurs aberrantes ?