Cours Math310

Première partie : statistique descriptive (Suite)

Paramètres de position et de dispersion

M. BEZOUI

13 décembre 2011

Plan de travail

Plan de travail

Médiane (définition)

Soit X, une variable statistique ayant k modalité ordonnée par ordre croissant : $x_1, x_2, ..., x_k$.

Définition

La médiane, noté M_e, est la valeur qui partage une série ordonnée, en deux parties d'effectif égale. En générale, M_e est la v.s, telle que, $F(M_e) = \frac{1}{2}$ ou, $N(Me) = \frac{n}{2}$.

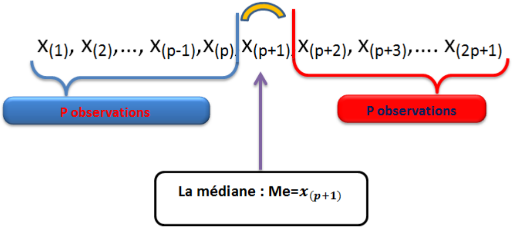
Soient les observations (données brutes) suivantes, préalablement ordonnées.

$$x_{(1)}, x_{(2)}, \dots, x_{(n)}$$

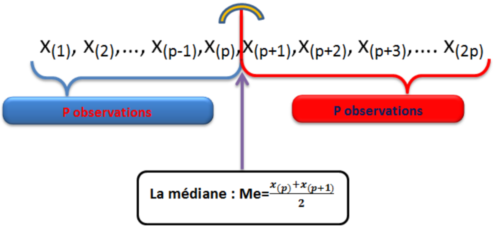
Dans la recherche de la médiane, on distigue deux cas :

- Nombre d'observation est impair,
- Nombre d'observation est pair.

Cas discret, n est impair ($n = 2p + 1$)



Cas discret, n est pair ($n = 2p$)



Exemple Numérique (cas discret)

Soient les observations suivantes :

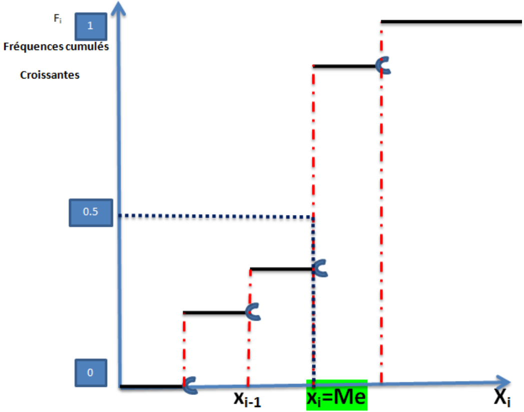
Données brutes $1, 4, 0, 1, 0, 1, 4, 3, 2, 1$

Observations ordonnées $0, 0, 1, 1, \mathbf{1}, \mathbf{1}, 2, 3, 4, 4$

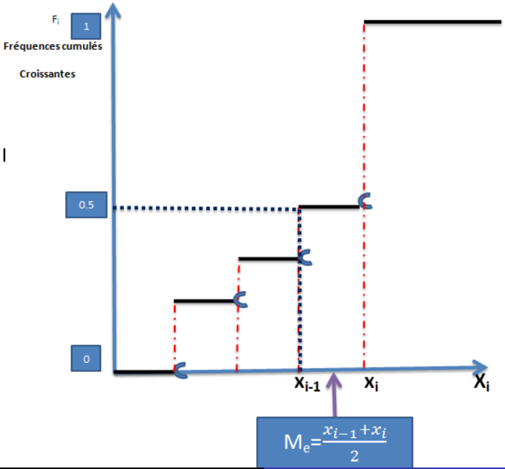
Es que n est pair ? Oui, $n = 10 = 5 \times 2 \Longrightarrow p = 5$

Trouver la médiane $M_e = \frac{x_p + x_{p+1}}{2} = \frac{1+1}{2} = 1$

Cas discret, Aucun pallier horizontal n'a $O.5$ comme ordonnée



Cas Discret, Un pallier horizontal a 0.5 comme ordonnée



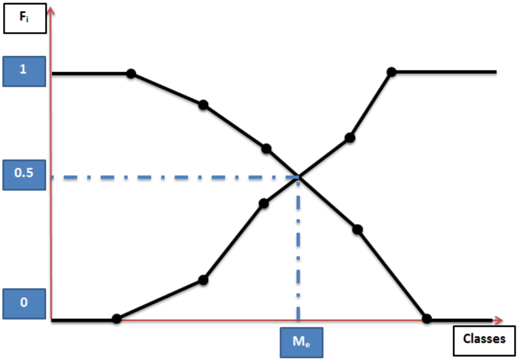
La médiane : Cas continu

$$M_e = e_{i-1} + a_i \frac{\frac{1}{2} - F(e_{i-1})}{F(e_i) - F(e_{i-1})}$$

Tels que :

- $[e_{i-1}, e_i[$: est la classe médiane,
- a_i est l'amplitude de la classe médiane $(a_i = e_i - e_{i-1})$

Graphiquement :



Plan de travail

La moyenne arithmétique

Noté $\overline{x}$, c'est la somme de toutes les observations divisée par n.

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{k} n_i x_i$$

Cas discret $\overline{x} = \frac{1}{n}\sum_{i=1}^{k} n_i x_i = \sum_{i=1}^{k} f_i x_i$

Cas continu $\overline{x} = \frac{1}{n}\sum_{i=1}^{k} n_i x_i = \sum_{i=1}^{k} f_i x_i$

x_i : Centres de classes.

k : Nombre de classes.

La moyenne arithmétique

Noté $\overline{x}$, c'est la somme de toutes les observations divisée par n.

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{k} n_i x_i$$

Cas discret $\overline{x} = \frac{1}{n}\sum_{i=1}^{k} n_i x_i = \sum_{i=1}^{k} f_i x_i$

Cas continu $\overline{x} = \frac{1}{n}\sum_{i=1}^{k} n_i x_i = \sum_{i=1}^{k} f_i x_i$

x_i : Centres de classes.

k : Nombre de classes.

La moyenne arithmétique

Noté $\overline{x}$, c'est la somme de toutes les observations divisée par n.

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{k} n_i x_i$$

Cas discret $\overline{x} = \frac{1}{n}\sum_{i=1}^{k} n_i x_i = \sum_{i=1}^{k} f_i x_i$

Cas continu $\overline{x} = \frac{1}{n}\sum_{i=1}^{k} n_i x_i = \sum_{i=1}^{k} f_i x_i$

x_i : Centres de classes.

k : Nombre de classes.

Propriétés de la moyenne

1. $\frac{1}{n}\sum_{i=1}^{k} n_i(x_i - \overline{x}) = 0$
2. Changement d'origine : Soit :$x_i' = x_i - x_0$, tel que x_0 : constante appelée origine.
   Alors la nouvelle moyenne sera : $\overline{x} = \overline{x'} + x_0$
3. Changement d'origine et d'echelle : $x_i' = \frac{x_i - x_0}{a}$
   - x_0 : origine (mode, médiane)
   - a : echelle (différence entre deux valeurs successives, ou l'amplitude unité dans le cas continu)

   La nouvelle moyenne sera alors : $\overline{x} = a\overline{x'} + x_0$

Propriétés de la moyenne

1. $\frac{1}{n}\sum_{i=1}^{k} n_i(x_i - \overline{x}) = 0$
2. Changement d'origine : Soit :$x'_i = x_i - x_0$, tel que x_0 : constante appelée origine.
   Alors la nouvelle moyenne sera : $\overline{x} = \overline{x'} + x_0$
3. Changement d'origine et d'echelle : $x'_i = \frac{x_i - x_0}{a}$
   - x_0 : origine (mode, médiane)
   - a : echelle (différence entre deux valeurs successives, ou l'amplitude unité dans le cas continu)

   La nouvelle moyenne sera alors : $\overline{x} = a\overline{x'} + x_0$

Propriétés de la moyenne

1. $\frac{1}{n}\sum_{i=1}^{k} n_i(x_i - \overline{x}) = 0$
2. Changement d'origine : Soit :$x_i' = x_i - x_0$, tel que x_0 : constante appelée origine.
   Alors la nouvelle moyenne sera : $\overline{x} = \overline{x'} + x_0$
3. Changement d'origine et d'echelle : $x_i' = \frac{x_i - x_0}{a}$
   - x_0 : origine (mode, médiane)
   - a : echelle (différence entre deux valeurs successives, ou l'amplitude unité dans le cas continu)

   La nouvelle moyenne sera alors : $\overline{x} = a\overline{x'} + x_0$

Propriétés de la moyenne

1. $\frac{1}{n}\sum_{i=1}^{k} n_i(x_i - \overline{x}) = 0$
2. Changement d'origine : Soit :$x_i' = x_i - x_0$, tel que x_0 : constante appelée origine.
   Alors la nouvelle moyenne sera : $\overline{x} = \overline{x'} + x_0$
3. Changement d'origine et d'echelle : $x_i' = \frac{x_i - x_0}{a}$
   - x_0 : origine (mode, médiane)
   - a : echelle (différence entre deux valeurs successives, ou l'amplitude unité dans le cas continu)

   La nouvelle moyenne sera alors : $\overline{x} = a\overline{x'} + x_0$

Propriétés de la moyenne

1. $\frac{1}{n}\sum_{i=1}^{k} n_i(x_i - \overline{x}) = 0$
2. Changement d'origine : Soit :$x_i' = x_i - x_0$, tel que x_0 : constante appelée origine.
   Alors la nouvelle moyenne sera : $\overline{x} = \overline{x'} + x_0$
3. Changement d'origine et d'echelle : $x_i' = \frac{x_i - x_0}{a}$
   - x_0 : origine (mode, médiane)
   - a : echelle (différence entre deux valeurs successives, ou l'amplitude unité dans le cas continu)

La nouvelle moyenne sera alors : $\overline{x} = a\overline{x'} + x_0$

Propriétés de la moyenne

1. $\frac{1}{n}\sum_{i=1}^{k} n_i(x_i - \overline{x}) = 0$
2. Changement d'origine : Soit :$x'_i = x_i - x_0$, tel que x_0 : constante appelée origine.
   Alors la nouvelle moyenne sera : $\overline{x} = \overline{x'} + x_0$
3. Changement d'origine et d'echelle : $x'_i = \frac{x_i - x_0}{a}$
   - x_0 : origine (mode, médiane)
   - a : echelle (différence entre deux valeurs successives, ou l'amplitude unité dans le cas continu)

   La nouvelle moyenne sera alors : $\overline{x} = a\overline{x'} + x_0$

Plan de travail

L'étendue

L'étendue d'une distribution est égale à la différence entre la plus grande et la plus petite valeur de la distribution :

$$e = max_{i=\overline{1,k}}(x_i) - min_{i=\overline{1,k}}(x_i)$$

L'étendue

L'étendue d'une distribution est égale à la différence entre la plus grande et la plus petite valeur de la distribution :

$$e = max_{i=\overline{1,k}}(x_i) - min_{i=\overline{1,k}}(x_i)$$

La variance, notée $V(x)$ est la moyenne du carré des écarts à la moyenne.

La variance

$$V(x) = \frac{1}{n}\sum_{i=1}^{k} n_i(x_i - \overline{x})^2 = \sum_{i=1}^{k} f_i(x_i - \overline{x})^2$$

L'écart type, noté σ_x est la racine carrée de la moyenne du carré des écarts à la moyenne, c'est à dire la racine carrée de la variance.

$$\sigma_x = \sqrt{V(x)} = \sqrt{\sum_{i=1}^{k}(x_i - \overline{x})^2}$$

Propriétés de la variance

1. $V(x) \geq 0$
2. $V(x) = \frac{1}{n}\sum_{i=1}^{k} n_i.x_i^2 - \overline{x}^2$
3. $V(x) = \sum_{i=1}^{k} f_i.x_i^2 - \overline{x}^2$
4. Soit $x_i' = \frac{x_i - b}{a} \Longrightarrow x_i = ax_i' + b$ alors
    1. $V(x) = a^2V(x') + b$
    2. $\sigma_x = |a|\sigma_{x'}$

Propriétés de la variance

1. $V(x) \geq 0$
2. $V(x) = \frac{1}{n}\sum_{i=1}^{k} n_i.x_i^2 - \overline{x}^2$
3. $V(x) = \sum_{i=1}^{k} f_i.x_i^2 - \overline{x}^2$
4. Soit $x_i' = \frac{x_i - b}{a} \Longrightarrow x_i = ax_i' + b$ alors
   1. $V(x) = a^2 V(x') + b$
   2. $\sigma_x = |a|\sigma_{x'}$

Propriétés de la variance

1. $V(x) \geq 0$
2. $V(x) = \frac{1}{n}\sum_{i=1}^{k} n_i.x_i^2 - \overline{x}^2$
3. $V(x) = \sum_{i=1}^{k} f_i.x_i^2 - \overline{x}^2$
4. Soit $x'_i = \frac{x_i - b}{a} \Longrightarrow x_i = ax'_i + b$ alors
   1. $V(x) = a^2 V(x') + b$
   2. $\sigma_x = |a|\sigma_{x'}$

Propriétés de la variance

1. $V(x) \geq 0$
2. $V(x) = \frac{1}{n} \sum_{i=1}^{k} n_i . x_i^2 - \overline{x}^2$
3. $V(x) = \sum_{i=1}^{k} f_i . x_i^2 - \overline{x}^2$
4. Soit $x_i' = \frac{x_i - b}{a} \Longrightarrow x_i = a x_i' + b$ alors
    1. $V(x) = a^2 V(x') + b$
    2. $\sigma_x = |a| \sigma_{x'}$

Propriétés de la variance

1. $V(x) \geq 0$
2. $V(x) = \frac{1}{n} \sum_{i=1}^{k} n_i.x_i^2 - \overline{x}^2$
3. $V(x) = \sum_{i=1}^{k} f_i.x_i^2 - \overline{x}^2$
4. Soit $x_i' = \frac{x_i - b}{a} \Longrightarrow x_i = ax_i' + b$ alors
    1. $V(x) = a^2 V(x') + b$
    2. $\sigma_x = |a|\sigma_{x'}$

Propriétés de la variance

1. $V(x) \geq 0$
2. $V(x) = \frac{1}{n}\sum_{i=1}^{k} n_i.x_i^2 - \overline{x}^2$
3. $V(x) = \sum_{i=1}^{k} f_i.x_i^2 - \overline{x}^2$
4. Soit $x_i' = \frac{x_i - b}{a} \Longrightarrow x_i = ax_i' + b$ alors
   1. $V(x) = a^2V(x') + b$
   2. $\sigma_x = |a|\sigma_{x'}$

Remarque

La série qui a un ecart type petit, est moins dispersée.

Les quartiles

Les quartiles d'ordre α, notés Q_α ; sont les valeurs de la v.s. qui sont solutions de l'equation $F(x) = \alpha$,

$$F(Q_1) = \frac{1}{4},\ F(Q_2) = F(M_e) = \frac{1}{2},\ F(Q_3) = \frac{3}{4}$$

$$N(Q_1) = \frac{1}{4}n,\ N(Q_2) = N(M_e) = \frac{1}{2}n,\ N(Q_3) = \frac{3}{4}n$$

Les quartiles

Les quartiles d'ordre α, notés Q_α ; sont les valeurs de la v.s. qui sont solutions de l'equation $F(x) = \alpha$,

$$F(Q_1) = \frac{1}{4},\ F(Q_2) = F(M_e) = \frac{1}{2},\ F(Q_3) = \frac{3}{4}$$

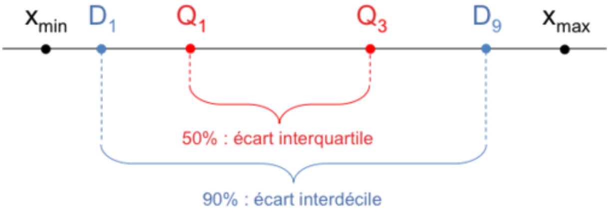
$$N(Q_1) = \frac{1}{4}n,\ N(Q_2) = N(M_e) = \frac{1}{2}n,\ N(Q_3) = \frac{3}{4}n$$

Les déciles

Les 9 déciles sont les nombres réels qui partagent l'étendue en dix intervalles de même effectif.

Intervalle Interqaurtiles

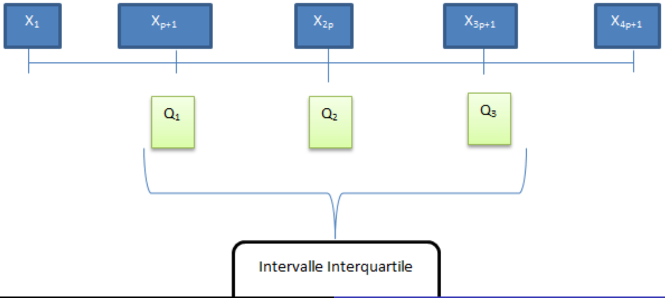
Intervalle qui contient 50% de l'effectif, laissant à gauche et à droite 25%



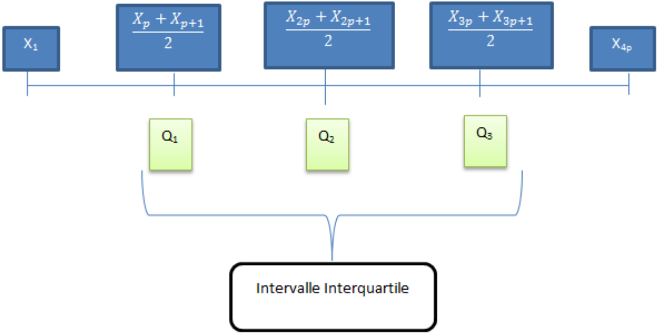
Pour cela, il faut déterminer Q_1, Q_3 avec la même méthode qu'on a utiliser pour trouver la médiane.

Exemple : Cas discret

Si on peut écrire n, sous la forme $n = 4p$ ou $n = 4p + 1$, alors on détermine la médiane, ensuite on détermine Q_1, Q_3, tel que Q_1 est la médiane de la partie gauche, ett Q_3, celle de la partie droite (comme dans la figure)



Intervalle Interquartiles



Moment simple d'ordre r

$$m_r = \frac{1}{n} \sum_{i=1} k n_i x_i^r$$

Moment centré d'ordre r

$$\mu_r = \frac{1}{n} \sum_{i=1} kn_i (x_i - \overline{x})^r$$

1. $\overline{x} = m_1$
2. $V(x) = m_2 - m_1^2$

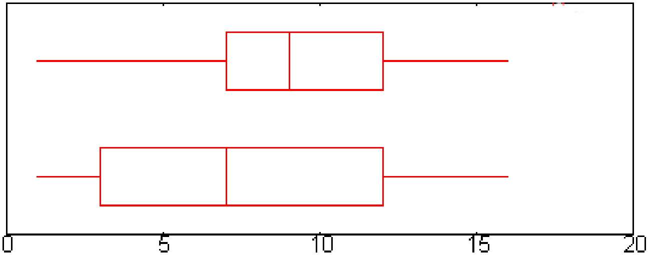
Plan de travail

Box Plot

Dans les représentations graphiques de données statistiques, la boîte à moustaches ou diagramme en boîte est un moyen rapide de figurer le profil essentiel d'une série statistique quantitative. Elle a été inventée en 1977 par John Tukey, mais peut faire l'objet de certains aménagements selon les utilisateurs. Son nom est la traduction de Box and Whiskers Plot.

Box Plot

La boîte à moustaches résume seulement quelques caractéristiques de position du caractère étudié (médiane, quartiles, minimum, maximum ou déciles). Ce diagramme est utilisé principalement pour comparer un même caractère dans deux populations de tailles différentes. Il s'agit de tracer un rectangle allant du premier quartile au troisième quartile et coupé par la médiane. Ce rectangle suffit pour le diagramme en boîte. On ajoute alors des segments aux extrémités menant jusqu'aux valeurs extrêmes, ou jusqu'aux premier et neuvième déciles (D1 / D9), voire aux 5e et 95e centiles. On parle alors de diagramme en boîte à moustaches ou de diagramme à pattes.



Comparaison de deux diagrammes en boîte à moustaches D_1 / D_9 avec

- pour la boîte inférieure : $Q1 = 3$, $M = 7$, $Q3=12$, $D_1 = 1$, $D_9 = 16$

- pour la boîte supérieure : $Q1 = 7$, $M = 9$, $Q3=12$, $D_1 = 1$, $D_9 = 16$

Merci pour votre attention