Cours Math310

Première partie : statistique descriptive (Suite)

Paramètres de dispersion et le BoxPlot

M. BEZOUI

3 janvier 2012

Plan de travail

1. Paramètres de dispersion
   - Les quartiles
   - Les déciles
   - Intervalle interquartiles
   - Moment simple d'ordre r
   - Moment centré d'ordre r
2. La boite à moustache ou Box-Plot
   - Présentation
   - Principe

Plan de travail

Dans le cous précédent

Dans le cours précédent

Etendue

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

Variance

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

Ecart type

$$\sigma_x = \sqrt{V(x)}$$

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 $F(x) = \alpha$,

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

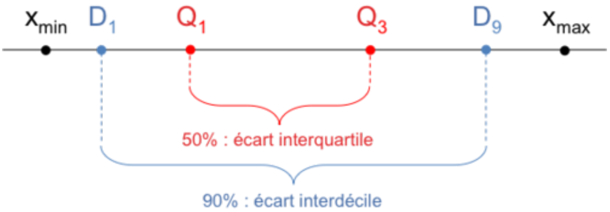
$$N(Q_1) = \frac{1}{4}n, \; N(Q_2) = N(M_e) = \frac{1}{2}n, \; N(Q_3) = \frac{3}{4}n$$

Les déciles

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

Intervalle Interqaurtiles

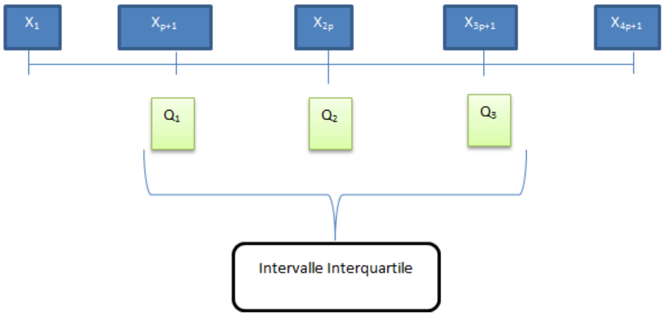
Intervalle qui contient 50% de l'effectif, laissant à gauche et à droite 25%



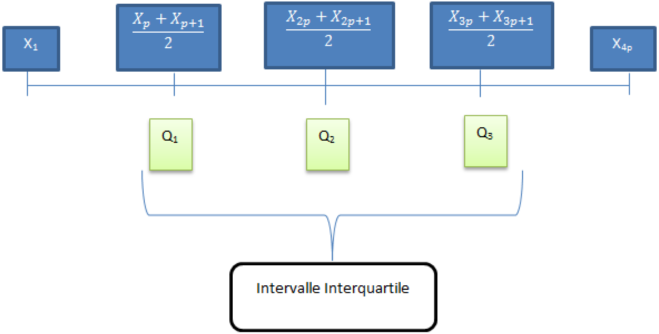
Pour cela, il faut déterminer Q_1, Q_3 avec la même méthode qu'on a utiliser pour trouver la médiane.

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, et 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

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

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

Plan de travail

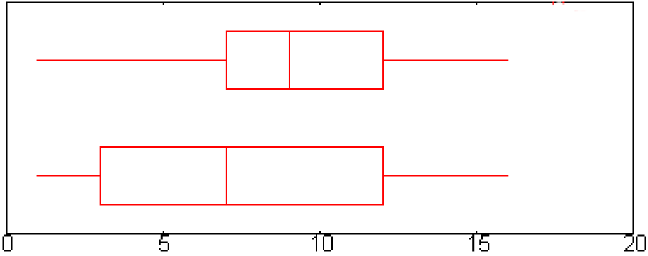
1. Paramètres de dispersion
   - Les quartiles
   - Les déciles
   - Intervalle interquartiles
   - Moment simple d'ordre r
   - Moment centré d'ordre r
2. La boite à moustache ou Box-Plot
   - Présentation
   - Principe

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, $D_1 = 1$, $D_9 = 16$

Exemple illustratif

Cosidérons la série statistique suivante :

Valeurs du caractère	Effectif	Effectif cumulé croissants $N_i \nearrow$
30	2	
45	3	
50	2	
60	2	
61	2	

Exemple illustratif (Suite)

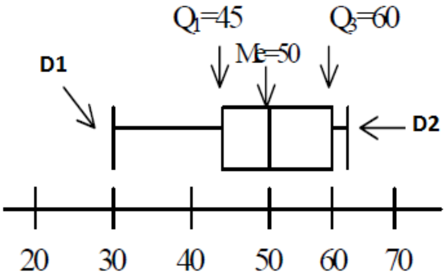
Cosidérons la série statistique suivante :

Valeurs du caractère	Effectif	Effectif cumulé croissants $N_i \nearrow$
30	2	2
45	3	5
50	2	7
60	2	9
61	2	11

Exemple illustratif (suite)

On trouve $Q_1 = 45$; $M_e = Q_1 = 50$; $Q_3 = 60$. D'autre part $D_1 = 30$; $D9 = 61$.

On peut alors représenter graphiquement ces résultats de la manière suivante :



Références:

- Eléments de statistique : 8183/22
- Exercices de statistiques descriptive 8183/28
- Cours de statistique descriptive 8183/30
- Statistique descriptive (cours et exercices) 8183/83
- Introduction à la statistique descriptive 8183/15