Chapitre 3

Caractéristiques de dispersion

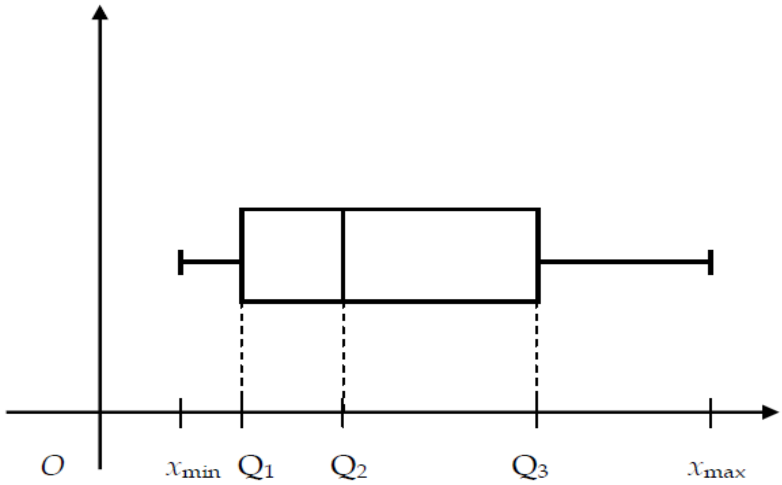
Compléments du cours

Prof. Mohamed El Merouani
Département de Statistique et Informatique
Faculté Polydisciplinaire de Tétouan

Boîte de Tuckey ou diagramme de Box & Wiskers

- Considérons le diagramme en boîte ci-dessous, qui est la version la plus simple de la boîte de Tuckey, appliquée à la variable X.

Boîte de Tuckey ou diagramme de Box & Wiskers



Boîte de Tuckey ou diagramme de Box & Wiskers

- On distingue sur ce schéma la « boîte de Tuckey » qui est le rectangle construit sur le premier quartile Q_1 et le troisième Q_3.

Exemple1: (cas de variable discrète "pondérée") :

- Calculer les quartiles Q_1, Q_2 et Q_3 de la série statistique suivante :

x_i	n_i	n_icc
3	3	3
4	7	10
8	30	40
10	20	60←
11	15	→75
20	25	100

N=100

1) $\frac{N}{4} = \frac{100}{4} = 25$

On cherche ce 25 entre les n_icc

$\Rightarrow$ il n'existe pas exactement mais 40 est la 1ère valeur qui dépasse ce 25

alors Q_1=8

5

2) $$\frac{N}{4} \times 2 = \frac{100}{2} = 50$$

$\Rightarrow$ ce 50 n'existe pas exactement parmi les n_icc, alors la 1ère valeur qui la dépasse est 60, donc

$\boxed{Q_2\text{=10=Mé}}$

3) $$\frac{N}{4} \times 3 = \frac{100}{4} \times 3 = 25 \times 3 = 75$$

Dans ce cas cette valeur existe exactement parmi les n_icc, alors

$$\boxed{Q_3 = \frac{11+20}{2} = 15{,}5}$$

Prof. Mohamed El Merouani

6

Exemple2 (cas de variable continue) :

Calculons Q_1, Q_2 et Q_3 de la distribution statistique suivante :

$[e_{i-1}, e_i[$	n_i	n_icc
[0, 10[	4	4
[10, 30[	8	12 *
[30, 35[	13	25
[35, 80[	5	30**
[80,100[	3	33
[100, 150[	7	40
N=	40	

Q_1= ? $\frac{N}{4} = \frac{40}{4} = 10$

cette valeur n'apparaît pas parmi les nicc, mais la 1ère valeur qui la dépasse est 12. D'où, on applique, dans la classe [10,30], la formule:

$$Q_1 = e_{i-1} + \frac{\frac{N}{4} - n_{i-1}cc}{n_i} a_i = 25$$

2) Q_2? $\frac{N}{4} \times 2 = \frac{40}{2} = 20$

$\Rightarrow$ Cette valeur n'existe pas non plus parmi les $n_i$cc, mais la 1ère valeur qui la dépasse est 25. Donc, on applique la formule dans la classe [30,25] :

$$Q_2 = Mé = e_{i-1} + \frac{\frac{N}{2} - n_{i-1}cc}{n_i} a_i$$

$$= 30 + \frac{20-12}{13} \times 5$$

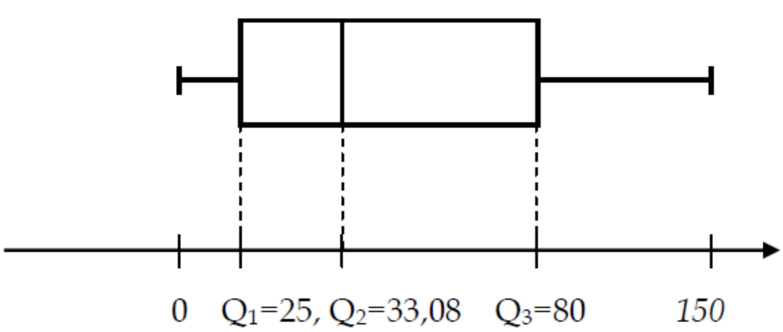
$$Q_2 \cong 33{,}08$$

3) Q_3? $\frac{N}{4}\times 3 = \frac{40\times 3}{4} = 30$

Cette fois-ci, cette valeur existe exactement parmi les $n_i$cc, la classe qui contient Q_3 est [35,80] .

Alors, on prend $Q_3 = e_i = 80$

Boîte de Tuckey ou diagramme de Box & Wiskers



- L'intervalle interquartile $=Q_3-Q_1=80-25=55$
- La dérivation quartile ou le semi-interquartile est:

$$\frac{Q_3-Q_1}{2}=\frac{55}{2}=27{,}5$$

- L'écart interquartile relatif est:

$$\frac{Q_3-Q_1}{Q_2}\cong\frac{55}{33{,}08}\approx 1{,}66$$

Exemple de calcul de la variance:

- Calculer la variance par la formule non-simplifiée et par la formule simplifiée, pour la série suivante :

$[e_{i-1}, e_i[$	n_i
[0, 10[	1
[10, 20[	2
[20, 30[	3
[30, 40[	4
	N=10

- La moyenne $\overline{X} = \frac{1}{N}\sum_{i=1}^{k} n_i c_i$
- Formule non-simplifiée de la variance

$$Var(X) = \frac{1}{N}\sum_{i=1}^{k} n_i \left(c_i - \overline{X}\right)^2$$

- Formule simplifiée de la variance

$$\mathrm{Var(X)} = \left(\frac{1}{N}\sum_{i=1}^{k} n_i c_i^2\right) - \overline{X}^2$$

Exemple de calcul de la variance::

$[e_{i-1}, e_i[$	n_i	c_i	c_i^2	$n_i c_i$	$c_i - \overline{X}$	$\left(c_i - \overline{X}\right)^2$	$n_i\left(c_i - \overline{X}\right)^2$	$n_i c_i^2$
[0, 10[	1	5	25	5	-20	400	400	25
[10, 20[	2	15	225	30	-10	100	200	450
[20, 30[	3	25	625	75	0	0	0	1875
[30, 40[	4	35	1225	140	10	100	400	4900
	N=10			250			1000	7250

- Alors, la moyenne est $\overline{X} = \frac{1}{N}\sum_{i=1}^{k} n_i c_i = \frac{250}{10} = 25$
- Donc, par la formule non simplifiée, la variance est:

$$\mathrm{Var(X)} = \frac{1000}{10} = 100$$

- Et par la formule simplifiée, elle sera:

$$Var(X) = 725 - (25)^2 = 725 - 625 = 100$$

- D'où, l'écart-type est

$$\sigma_X = \sqrt{\mathrm{Var(X)}} = \sqrt{100} = 10$$

Pour l'exemple de calcul de la variance:

- **Le coefficient de variation**

$$\mathrm{C_V} = \frac{\sigma(\mathrm{X})}{\overline{\mathrm{X}}} = \frac{10}{25} = 0,4$$

- **Soit donc un coefficient de variation de 40% supérieur à 30%**
- **Donc, la série étudiée et non-homogène!**