

Compléments concernant les graphiques

IUT Bordeaux Montesquieu - Département Tech De Co - Bernard Andruccioli - Septembre 2005

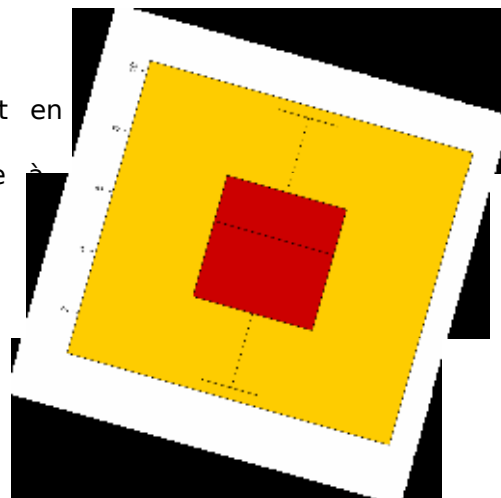
Ce document complète les informations disponibles en ligne (<http://andruccioli.u-bordeaux4.fr>). Il s'agit d'un support de cours, et d'un complément qui ne peut en aucun cas se substituer au cours.

La boîte à moustaches de Tukey

Parmi les nombreuses représentations graphiques, peu mettent en évidence un ou plusieurs paramètres de tendance centrale, et un paramètre de dispersion. C'est le cas de la boxplot ou boîte à moustaches de Tukey qui permet, avec un minimum d'habitude, d'appréhender visuellement l'ordre de grandeur d'une série ainsi que sa dispersion.

Remarque : un article remarquable de Monique Le Guen se trouve à cette adresse :

<http://matisse.univ-paris1.fr/leguen/leguen2001b.pdf>



Présentation

Objectif : représenter schématiquement une distribution statistique en intégrant des paramètres de tendance centrale et de dispersion ; la lecture de ce type de graphique devient complète et les comparaisons entre plusieurs séries sont visuellement possibles.

Il en existe plusieurs versions, mais une telle boîte comprend le plus souvent :

- Une échelle des valeurs présente sur l'axe vertical

- La valeur du premier quartile (Q1) qui est le trait inférieur de la boîte.

- La valeur du troisième (Q3) qui est le trait supérieur de la boîte

- La valeur de second quartile (Q2 qui est donc la médiane) représentée par un trait horizontal au sein de la boîte.

- Les moustaches inférieure et supérieure représentées par des petits rectangles verticaux de chaque côté de la boîte.

- La moyennes est parfois présente et représentée par un +.

- Les valeurs maximum et minimum sont parfois représentées (cercle ou *) de même que certaines valeurs atypiques.

Mise en oeuvre

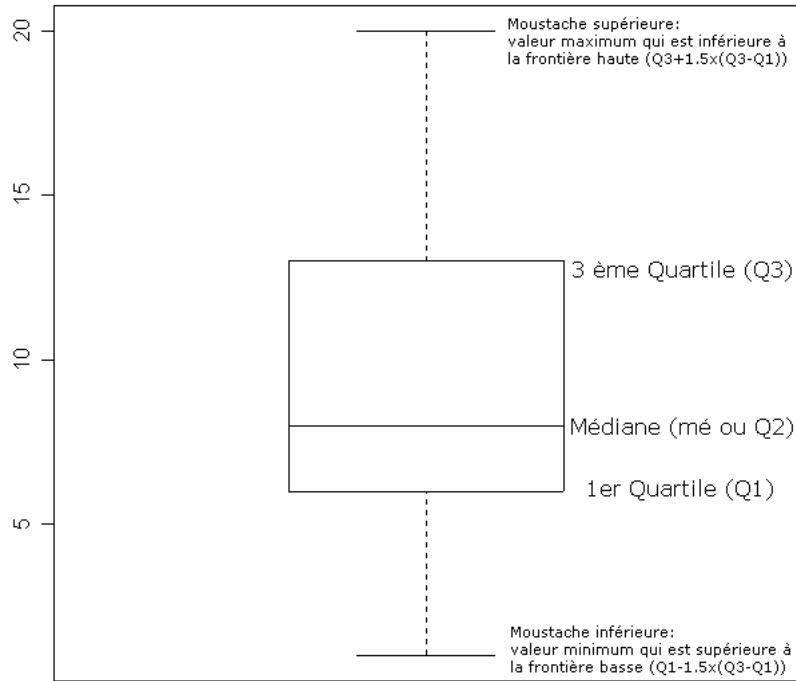
Nous nous en tiendrons aux premiers éléments mentionnés ci-dessus.

Les valeurs des moustaches inférieure et supérieure sont calculées ainsi :

Moustache inférieure : valeur de la série immédiatement supérieure à la frontière basse, avec la frontière basse qui est égale à $Q1 - 1,5 \times (Q3 - Q1)$

Moustache supérieure : valeur de la série immédiatement inférieure à la frontière haute, avec la frontière haute égale à : $Q3 + 1,5 \times (Q3 - Q1)$

Ces précautions évitent de se trouver avec des valeurs négatives, pour la moustache inférieure, ou beaucoup trop grandes (en dehors de la série) pour la moustache supérieure. D'où :



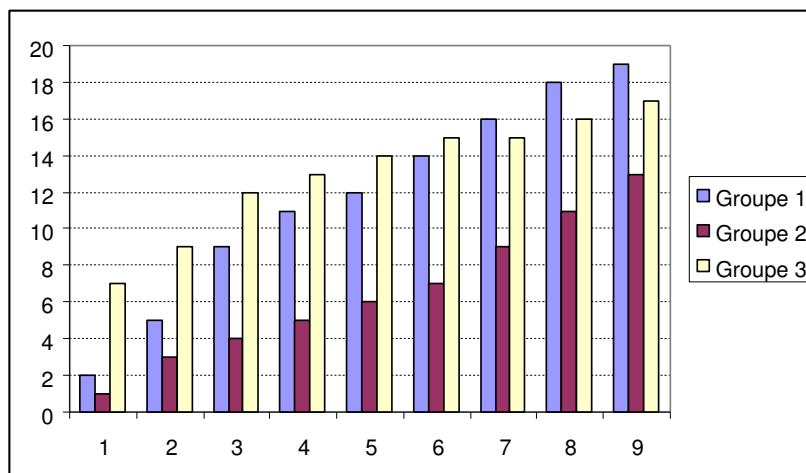
La boîte à moustaches de TUKEY tracée avec le logiciel R.

Intérêt (exemple)

Le tracé d'un tel graphique oblige donc à calculer certains indicateurs, mais avec un peu d'habitude sa lecture est bien plus pertinente que des graphiques habituels. Voyons l'exemple suivant qui reprend des notes obtenues par trois groupes d'étudiants à un contrôle d'informatique :

Groupe 1	Groupe 2	Groupe 3
2	1	7
5	3	9
9	4	12
11	5	13
12	6	14
14	7	15
16	9	15
18	11	16
19	13	17

Le premier réflexe de représentation (accentué par les tableurs) est celui-ci :



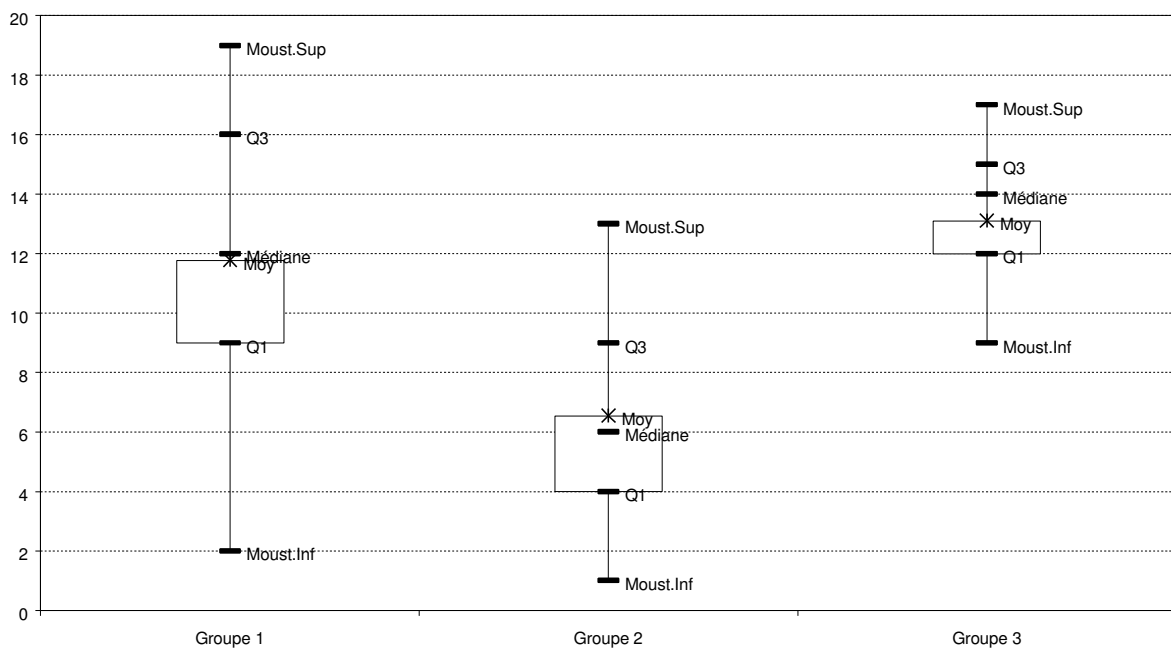
On y voit l'ensemble des notes, mais si nous souhaitons comparer les résultats de ces groupes nous ne sommes guère plus avancés.

Calculons les indicateurs nécessaires : (Une formule par cellule sur un tableue, mais on peut recopier vers la droite le plus souvent)

	Groupe 1	Groupe 2	Groupe 3
Q1	9	4	12
Moust.Inf	2	1	9
Médiane	12	6	14
Moust.Sup	19	13	17
Q3	16	9	15
Moy	11.8	6.6	13.1

Et créons ce mode de représentation (pas si simple avec Excel...)

Boxplot ou boîte à moustaches de Tukey



D'un simple coup d'œil on visualise le niveau moyen des groupes, la répartition des notes...la comparaison est donc aisée.

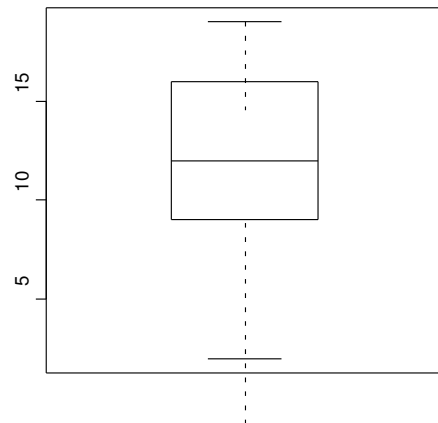
Sur Excel obtenir un tel graphique est assez fastidieux (pas de modèle prédéfini). L'utilisation de logiciels statistiques se comprend facilement.

Utilisation de R

Exemple avec le logiciel R : Bien qu'il fonctionne en ligne de commandes, on gagne un temps précieux ;

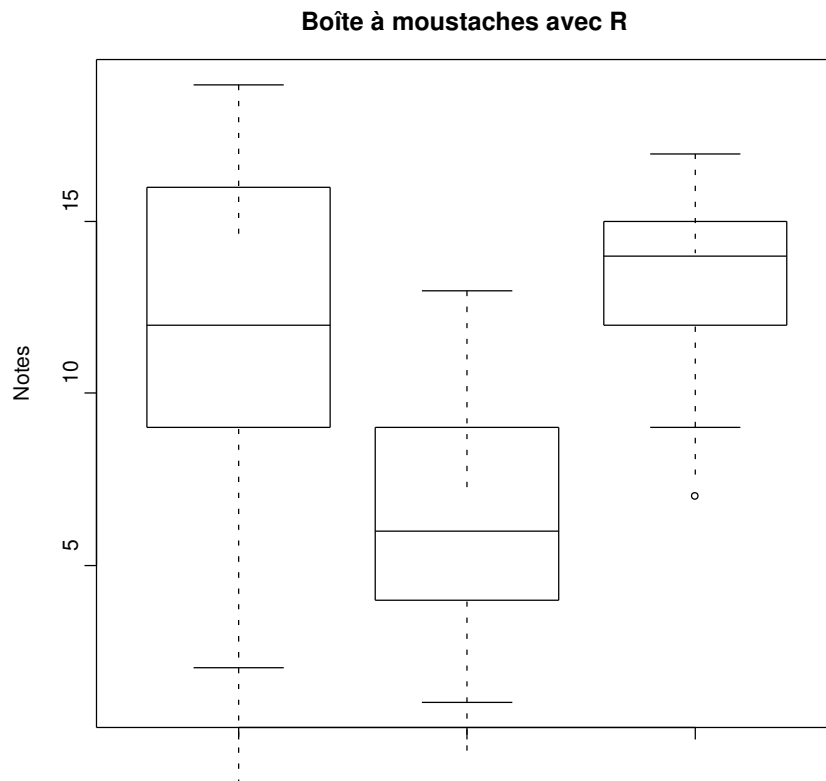
Saisie des valeurs :

```
> groupe1<-  
c(2,5,9,11,12,14,16,18,19)  
> groupe2<- c(1,3,4,5,6,7,9,11,13)  
> groupe3<-  
c(7,9,12,13,14,15,15,16,17)  
Calcul d'indicateurs essentiels (pas  
obligatoire pour le tracé des graphiques :  
(Ci-contre le groupe 1)  
> summary(groupe1)  
  Min. 1st Qu.  Median    Mean 3rd  
Qu.    Max.        
 2.00   9.00   12.00   11.78  
16.00  19.00  
Tracé de la boxplot pour le groupe 1 :  
> boxplot(groupe1)
```



Avec les trois groupes :

```
>boxplot(groupe1, groupe2, groupe3, main="Boîte à moustaches avec R", ylab="Notes")
```



<http://andruccioli.u-bordeaux4.fr> - Bernard.Andruccioli@u-bordeaux4.fr