Corrigé exercice 1 : statistique descriptive

Objectif : se donner une idée de la répartition des données

1. Pré-requis
   1. orthographe
   2. français

# Exemples non corrigés :

Quantiles (Q)

Il indique combien de pourcentage des observations qui se trouve en dessous de la valeur indiqué.

* Exemple : il y 25 % du chiffre observé qui sont inférieurs de 13.40335

1. Sources des données
   1. préciser la source + la date de recensement des données
   2. préciser ce que représentent les données (avec les unités éventuelles)

Dans R : Données – importer des données – depuis un fichier excel – cliquer sur éditer ou visualiser

# Exemple sur les débits de serveurs :

*Source : ?? + janvier 2012*

*Chaque donnée représente le débit d’un serveur iesn en mégabits/sec.*

1. Indicateurs statistiques :
   1. moyenne – écart-type – médiane – quartiles – iqr – min – max - N
   2. interprétation contextuelle

Dans R : Statistiques – résumé – statistiques descriptives – onglet « statistique » cocher coefficient de variation

*Résultats :*

|  |  |  |  |
| --- | --- | --- | --- |
| Moyenne | 14.11 | Percentiles |  |
| Ecart-type | 1.36 | 0% | 4.75 |
| IQR | 1.27 | 25% | 13.4 |
| Coefficient variation | 0.1 | 50% | 14.06 |
| N | 307/307 | 75% | 14.7 |

La moyenne des débits de serveurs observés est de 14.11 Mb/sec. L’écart-type est de 1.35 Mb/sec avec un coefficient de variation de 10%, ce qui est très petit et indique que les données sont assez rassemblées autour de la moyenne. Aucune donnée en dessous de 4.75 Mb/sec.

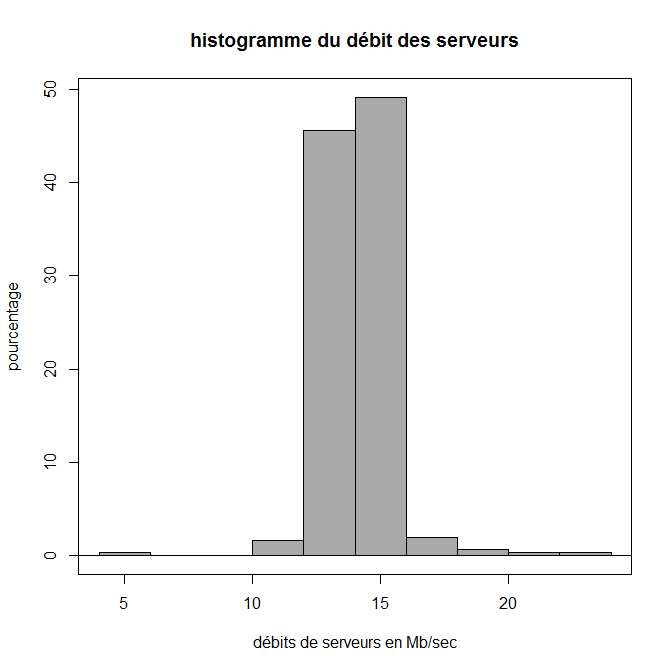
Quartile 1 vaut 13.4, ce qui indique que seulement 25% des données sont inférieures à 13.4Mb/sec.

Le quartile 2 = médiane vaut 14.05 Mb/sec. Ceci indique que 50% sont inférieures à cette valeur. La médiane est proche de la moyenne ainsi nous pouvons supposer que les données sont réparties de manière assez symétriques autour de la moyenne.

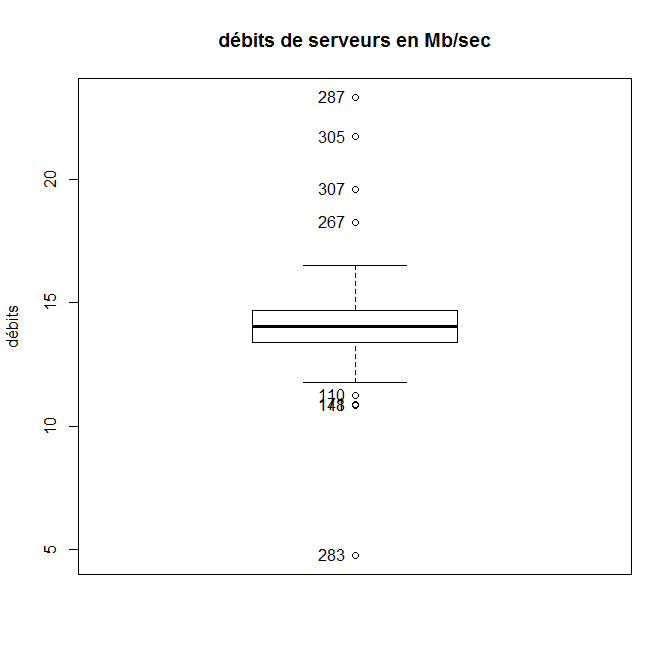
L’iqr est très petit. Ainsi, les données sont très rassemblées autour de la moyenne.

1. Graphiques pour des données numériques continues
   1. Histogramme : attention aux libellés des axes + lecture d’une barre de l’histogramme
   2. Box plot : lecture de la box plot.

Dans R : Graphes – histogramme – choisir débits + options. Pour le mettre dans word, cliquer sur le menu fichier du graphique – copier vers le presse papier – bitmap –coller dans word



*Lecture : environ 45% des débits de serveurs observés se trouvent entre 12 et 14Mb/sec.*



signification box plot

les valeurs au dessus et au dessous des moustaches sont des valeurs dites abérrante.