

Auteur : Pierre Lapôtre, Emmanuel Ostenne et Martijn van Brugge

Les tableurs, calculatrices et logiciels de calcul employés au lycée contiennent des fonctions qui calculent les premier et troisième quartiles d'une série statistique. Ces fonctions sont basées sur des définitions de Q_1 et Q_3 différentes des définitions des programmes français d'enseignement. Les résultats qu'elles donnent sont donc faux pour nous. Nous allons établir des algorithmes corrects exécutables soit sur le tableur « Calc », soit sur une calculatrice programmable, soit sur « scilab pour les lycées », « Xcas » ou « Javascript ».

Est donnée une série statistique

$$A := (a_1, \dots, a_n) \text{ de longueur } n \geq 2.$$

Première partie : tri de la série statistique

1 - Commençons par ranger A dans l'ordre croissant :

Algorithme 1 : Tri d'une série statistique

Entrées :

A : suite de nombres réels (série statistique)

Sorties :

B : suite obtenue en ordonnant A

début

 | Trier A

 | Afficher B

fin

« Trier » désigne dans l'algorithme 1 une fonction pré-programmée du système utilisé.

Comment faut-il modifier l'algorithme précédent si « Trier » trie dans l'ordre décroissant ?

Indication : Si $x \leq y$, $-x \geq -y$, $x = -(-x)$ et $y = -(-y)$.

Deuxième partie : calcul de la médiane

Nous allons programmer le calcul de la médiane m , bien que tous les systèmes pré-programmés donnent une réponse correcte à ce problème.

2 - Écrire en langage naturel un algorithme de calcul de m .

3 - Traduire cet algorithme en un algorithme exécutable par le système utilisé en classe.

4 - Pour tester cet algorithme,

4.a - engendrer une suite A de 1000000 de nombres au hasard dans l'intervalle $[0,1]$ (ou une suite plus courte, mais aussi longue que possible, si votre système ne le permet pas) ;

4.b - indiquer sans calcul quelle valeur remarquable devrait être très proche de m dans ce cas ;

4.c - vérifier votre pronostic en effectuant les calculs.

Troisième partie : calcul des premier et troisième quartiles

Dans cette partie, on appellera q et r le quotient et le reste de la division euclidienne de n par 4. On rappelle que

Le premier quartile Q_1 est le plus petit élément x de la série statistique tel qu'au moins 25% des données soient inférieures ou égales à x .

Le troisième quartile Q_3 est le plus petit élément y de la série statistique tel qu'au moins 75% des données soient inférieures ou égales à y .

5.a - On suppose que $r = 0$, c'est à dire que n est un multiple de 4. Exprimer Q_1 et Q_3 comme éléments convenablement choisis de B .

5.b - Même question si l'on suppose que $r = 1$.

Il resterait à examiner les cas $r = 2$ et $r = 3$. *On admettra le tableau récapitulatif suivant :*

| | | | | |
|-------|----------|------------|------------|------------|
| r | 0 | 1 | 2 | 3 |
| Q_1 | b_q | b_{q+1} | b_{q+1} | b_{q+1} |
| Q_3 | b_{3q} | b_{3q+1} | b_{3q+2} | b_{3q+3} |

6 - Écrire en langage naturel un algorithme de calcul de Q_1 et de Q_3 .

7 - Prendre connaissance de la version de cet algorithme qui est exécutable par le système utilisé en classe.

8 - Pour tester cet algorithme,

8.a - engendrer une suite A de 1000000 de nombres au hasard dans l'intervalle $[0,1]$ (cf. question **4.a**) ;

8.b - indiquer sans calcul quelles valeurs remarquables devraient être très proches de Q_1 et de Q_3 respectivement ;

8.c - vérifier votre pronostic en effectuant les calculs.