

**EPITECH**  
**Année 2011-2012 Rattrapage**  
**Projet 212analyse**

## **1 Objectif**

Une entreprise sur Internet veut connaître les habitudes de navigation de ses clients. L'entrepreneur souhaite analyser les données obtenues via un sondage réalisé en ligne à l'aide de méthodes statistiques. Il souhaiterait aussi savoir à partir de quelle taille les échantillons sont représentatifs de l'ensemble de la population.

## **2 Modélisation**

Le but de ce sujet est d'analyser des observations effectuées sur une population donnée, puis d'étudier les distributions d'échantillonnage des moyennes.

On dispose d'un fichier récapitulant les observations concernant le temps passé sur le site. Pour ces observations de  $N$  individus, on calculera dans un premier temps :

1. la moyenne arithmétique,
2. la moyenne géométrique,
3. la moyenne quadratique,
4. la médiane,
5. l'écart-type.

Dans un second temps, on considèrera la distribution d'échantillons de taille  $n$  ( $1 \leq n \leq N$ ). Pour cette distribution, on calculera :

1. le nombre de ces échantillons,
2. la moyenne arithmétique des temps relevés pour chaque échantillon (à ne pas afficher !).  
On affichera la moyenne la plus basse et la moyenne la plus haute.
3. la moyenne arithmétique de toutes les moyennes obtenue,
4. l'écart-type de toutes les moyennes obtenues.

## **3 Fichier de données**

Les observations sont stockées dans un fichier. Sur la première ligne du fichier se trouve le nombre total d'individus  $N$ . Sur chacune des  $N$  lignes suivantes, il y a le temps moyen passé par l'individu sur le site de l'entreprise ce dernier mois, en secondes.

## **4 Le logiciel**

Le ramassage se fait par SVN.

Le nom du dépôt est *212analyse-rattrapage-2015-LOGIN*.

Nom de l'exécutable : 212analyse

Exemple de lancer :

> ./212analyse observ.txt n

En entrée: le fichier de données (*observ.txt*), et la taille des échantillons ( $n$ ).

En sortie : la liste des 10 valeurs demandées arrondies à 2 chiffres après la virgule.

## 5 Remarque

Essayez de ne pas stocker tous les échantillons possibles. En effet, pour  $N$  suffisamment grand, vous auriez des tableaux de dimensions très élevées, ce qui est pénalisant en mémoire. Réalisez vos calculs de moyenne et d'écart-type des échantillons en stockant un minimum de valeurs.  
A vous d'utiliser les bonnes formules.

## 6 Moulinette

### Attention !

Le projet sera corrigé par une moulinette.  
Il est impératif de respecter scrupuleusement l'affichage de l'exemple.

Vous DEVEZ rendre un makefile contenant TOUTES les règles suivantes :  
\$(NAME), fclean, clean et all

Votre makefile devra générer via la règle all ou \$(NAME) votre exécutable, nommé 212analyse. Ceci est obligatoire car la première action de la moulinette sera de supprimer votre exécutable, s'il existe.

Enfin, comme vous le savez sûrement déjà, un programme s'exécutant sans erreur aucune est un programme qui retourne 0.  
Etant donné que les tests que nous effectuerons seront toujours valides, votre programme devra donc retourner 0 (EXIT\_SUCCESS).  
Si ce n'est pas le cas, vous vous verrez attribuer la note de 1.5.

N'oubliez pas d'insérer un shebang au début de votre exécutable si cela est nécessaire (ex : #!/bin/sh).

## 7 Exemple

```
> ./212analyse observ.txt 4
moyenne arithmetique : 697.50
moyenne geometrique : 645.00
moyenne quadratique : 705.32
mediane : 722.00
ecart-type : 62.15
nombre d'echantillons de taille 4 : 12356.00
moyenne la plus basse : 435.12
moyenne la plus haute : 1098.94
moyenne arithmetique des moyennes : 697.50
ecart-type des moyennes : 38.42
```