

**EPITECH**  
**PROBABILITES ET STATISTIQUES**  
**Année 2011-2012**  
**Mini-projet 209sondage**

## 1 Objectif

Une petite société possède un parc informatique de PC. Certains PC deviennent obsolètes alors que d'autres viennent d'être acquis. L'ingénieur système établit la liste des capacités mémoire de chaque machine. Il met au point un logiciel qui calcule les statistiques correspondantes.

## 2 Modélisation

Le but de ce sujet est de retrouver informatiquement les formules reliant paramètre et statistique pour l'étude d'une distribution échantillonnale des moyennes.

On dispose d'une liste de  $N$  capacités mémoires d'ordinateurs. Dans un premier temps, on établira les paramètres de la population (les capacités mémoires) qui sont :

- la moyenne  $\mu$  des capacités de la population,
- l'écart-type  $\sigma$  de la population.

Dans un deuxième temps, on considère la distribution échantillonnale d'échantillons de taille  $n$  ( $1 \leq n \leq N$ ). On commence par calculer le nombre de tels échantillons. Ensuite, pour chaque échantillon, on calcule la moyenne des capacités. Il ne reste plus alors qu'à calculer la moyenne  $\mu_{\overline{X}}$  et l'écart-type  $\sigma_{\overline{X}}$  de toutes les moyennes obtenues. Finalement, on demande de vérifier que  $\mu_{\overline{X}}$  et  $\mu$  et que  $\sigma_{\overline{X}}$  et  $\sigma$  sont reliés par les formules suivantes :

$$\mu_{\overline{X}} = \mu$$
$$\sigma_{\overline{X}}^2 = \frac{\sigma^2}{n} \cdot \frac{N-n}{N-1}$$

Les différentes quantités qu'affichera votre programme sont données dans les exemples ci-après.

## 3 Le logiciel

Le ramassage se fait par svn. Vous recevrez un mail, lorsque les dépôts seront ouverts. Si vous avez des questions sur le rendu, lisez la documentation **doc\_user\_kscm.pdf**, sinon demandez au labo Koala.

Nom de l'exécutable : 209sondage

Exemple de lancer :

> **209sondage** 209test1.txt n

La taille n des l'échantillons est passée en paramètre. Elle varie entre 1 et le nombre total d'ordinateurs N.

Le fichier contient en première ligne le nombre de valeurs à lire. Sur les lignes suivantes, une capacité mémoire par ligne.

## 4 Remarque

Essayez de ne pas stocker tous les échantillons possibles dans une table. En effet, pour N aux alentours de 20, vous auriez des tables de dimension très élevées ce qui est pénalisant en mémoire. Réalisez vos calculs de moyenne et d'écart-type au fur et à mesure du calcul des échantillons. A vous d'utiliser les bonnes formules...

## 5 Questions

- 1° Quel est l'avantage des méthodes d'échantillonnage aléatoires ?
- 2° Comment nomme-t-on la valeur inconnue d'une population à estimer ?
- 3° Comment s'appelle la valeur calculée sur un échantillon destinée à estimer le paramètre d'une population ?
- 4° Donner des exemples de paramètres de population à estimer.
- 5° Qu'appelle-t-on distribution d'échantillonnage ?

## 6 Exemples

Population : 209test1.txt

Nombre d'individus : 20

Moyenne m : 697

Ecart-type s : 413,39

Echantillons de taille : 6

Nombre d'échantillons : 38760

Moyenne : 697

Ecart-type : 144,868

Ecart-type calculé d'après s : 144,868

Population : 209test2.txt

Nombre d'individus : 15

Moyenne m : 1030,67

Ecart-type s : 411,817

Echantillons de taille : 4  
Nombre d'échantillons : 1365  
Moyenne : 1030,67  
Ecart-type : 182,518  
Ecart-type calculé d'après s : 182,518

Population : 209test3.txt  
Nombre d'individus : 10  
Moyenne : 1276  
Ecart-type : 575,486

Echantillons de taille : 3  
Nombre d'échantillons : 120  
Moyenne m : 1276  
Ecart-type s : 293,023  
Ecart-type calculé d'après s : 293,023