

Statistiques élémentaires - Indicateurs de position (1^{ère} partie)

- Préambule :**
0. (Rappel) Ouvrir une session en tapant **1L2** comme nom d'utilisateur – mot de passe « **jml** ».
 1. Ouvrir le classeur Excel à votre nom dans le dossier de **1L2_Ganesh / 1L2_2005_G1 ou G2**.
 2. Insérer une nouvelle feuille (Menu **Insertion / Feuille**) et la placer *après* les précédentes.
 3. Renommer cette feuille **TD6 Stat.1** comme pour les TD précédents.
 4. On prendra soin de rédiger les réponses aux questions dans la feuille de calcul en utilisant si nécessaire la commande de *fusion des cases* pour le texte.

I - Dans l'énoncé qui suit, tous les mots en caractère gras font partie du vocabulaire de base des statistiques élémentaires. Les définitions de ces mots se déduisent facilement de l'exemple auquel ils sont attachés ou du sens courant.

Énoncé du problème : Un fournisseur de chaussures de sport se propose de faire une rapide étude de marché pour déterminer quelles sont les tailles qui sont le plus demandées dans les lycées du voisinage.

Pour cela il prélève un **échantillon** dans la **population** constituée par les élèves du lycée le plus proche. On suppose que l'échantillon a un **effectif** de 50 élèves choisis de manière à rendre l'échantillon le plus représentatif possible, et il s'intéresse donc au « **caractère** » défini par la pointure des élèves pris dans cet échantillon ; il relève alors la **valeur** de ce caractère (en nombres entiers de cm) pour chaque individu.

Il obtient ainsi une **série** statistique brute (non classée) de 50 **nombres** entiers de 36 à 45 (inclus).

Il doit alors calculer la **fréquence** de chacune des pointures relevées, puis pour visualiser les résultats de son étude il doit construire un **histogramme des fréquences** trouvées pour chacune des « **classes** » définies par les différentes pointures.

Il va ainsi d'abord déterminer le **mode** de cette série (c'est à dire la pointure dont la fréquence est la plus élevée), ainsi que la **moyenne** et la **médiane** (c'est à dire la valeur de la pointure correspondant à 50% de l'effectif classé).

D'autres paramètres seront nécessaires pour affiner cette étude, et ils feront l'objet des prochains TD.

La Série brute relevée à partir de l'échantillon est donnée ci-dessous :

{43, 42, 40, 38, 41, 40, 41, 37, 39, 39, 42, 45, 40, 45, 45, 40, 42, 40, 43, 42, 43, 41, 41, 44, 45,
41, 39, 37, 42, 45, 41, 38, 43, 36, 38, 39, 40, 38, 36, 39, 45, 40, 43, 39, 44, 44, 43, 40, 38, 37 }

1°) Construire le tableau suivant dans la feuille du tableur .

Pour cela il faut d'abord compter manuellement l'effectif n_i de chaque classe x_i puis en déduire les fréquences f_i à l'aide de formules appropriées que l'on étendra sur toute la ligne.

Rappel des formules de calcul de la MOYENNE \bar{x} : on pose $N = \sum n_i = \text{effectif total}$. Ainsi on obtient :

$$(1) \bar{x} = \frac{\sum x_i}{N} \quad \text{ou} \quad (2) \bar{x} = \frac{\sum x_i n_i}{\sum n_i} \quad \text{ou} \quad (3) \bar{x} = \sum x_i f_i \quad \text{avec } f_i = \frac{n_i}{N} \dots$$

Classes : x_i	36	37	38	39	40	41	42	43	44	45	Totaux
Effectifs : n_i											N
Fréquences : f_i (%)											100 %
Produits $x_i f_i$											\bar{x}

Paramètres de position :

Mode	Moyenne	Médiane	1^{er} Quartile	3^e Quartile

2°) Construire l'**histogramme des fréquences** en indiquant les valeurs x_i des classes sur l'axe des abscisses.

II – Construction automatique aléatoire d'un échantillon et calcul automatique des effectifs puis des fréquences.

1°) Pour engendrer une série de nombres au hasard on peut utiliser une fonction définie dans le tableur nommée **ALEA()**.

Cette fonction qui n'a pas d'argument retourne automatiquement un nombre décimal à 7 chiffres compris entre 0 et 1 (exclus). Pour obtenir un nombre *entier* au hasard entre 36 et 45 il suffit donc de le multiplier par 10 puis d'éliminer la partie décimale et d'ajouter 36 au résultat : on écrit donc dans la première cellule **=ENT(10 * ALEA()) + 36** puis on recopie cette formule autant de fois qu'on le désire (par exemple sur 200 lignes d'une même colonne).

Comme la fonction ALEA() retourne un nombre à chaque fois que l'on valide une formule, il est nécessaire de configurer la feuille en mode « **calcul sur ordre** » (menu **Fichier / Préférences / OPTIONS / CALCUL**). Puis appuyer sur la touche **F9** pour le relancer autant de fois que l'on veut (ce qui évidemment modifie les résultats chaque fois).

2°) Donner un nom (« *série* ») à la colonne des nombres ainsi obtenus. [Menu : **Insertion / Nom**]. On peut alors faire calculer la Moyenne et la Médiane de cette série brute par les commandes **=MOYENNE(série)** et **=MEDIANE(série)**. (attention à ne pas oublier le signe = devant la commande).

3°) Pour déterminer le nombre d'occurrences d'un nombre donné dans une série brute, on utilise la fonction suivante :

=NB.SI(série ; valeur) qui retourne le nombre d'occurrences de la valeur x_i indiquée par une référence de cellule, dans la série indiquée. Pour utiliser efficacement cette fonction on construit une colonne contenant les valeurs x_i des classes, dont les éléments serviront de référence pour les valeurs.

3°) On complète le tableau obtenu avec les **fréquences**, puis on construit l'**histogramme** correspondant.

4°) En appuyant sur la touche **F9** on voit alors se modifier automatiquement le tableau des valeurs des paramètres de position et l'histogramme associé.