

TP info : Population des communes françaises et loi de Benford

Date

A rendre à l'issue de la séance : un fichier excel contenant votre travail (envoyé par mail avant la fin de la séance ou sur clé USB). Pour jeudi, me rendre un compte rendu par binôme (manuscrit ou tapé). Ce sera noté, même coeff qu'une interro de cours.

1 Introduction, position du problème

Le but de ce TP est de récupérer des bases de données existantes de diverses sources, par exemple la population de communes française, la superficie des pays du monde, la longueur des fleuves, etc ...

Cela nous donne plein de nombres, pour lesquels on ne s'intéresse qu'à la répartition du premier chiffre (le chiffre le plus à gauche dans l'écriture en base 10).

On donne un exemple ci-dessous :

Exercice 1 : Remplir les cases manquantes. En particulier, donner l'effectif total et expliquer le calcul de la fréquence d'occurrence.

Valeur	10	16	9	356	875	456	124	356	852	445	331	907
Premier chiffre	1	1	9	3	8	4	1	3	8	4	3	9
Premier chiffre			1	2	3	4	5	6	7	8	9	
Nombre d'occurrence dans les données			3	0	3	2	0	0				
Fréquence d'occurrence												

Question 1.1. *A priori, sur une large collection de nombres, y-a-t-il plus de chance que 1 apparaisse plus fréquemment comme premier chiffre que 9 ? Quelle serait donc la répartition des fréquences pour ce premier chiffre ?*

2 Exemple : population des communes françaises

Sur le site de l'INSEE, récupérer le fichier du recensement de 2013 des communes de l'Isère. <http://www.insee.fr/fr/ppp/bases-de-donnees/recensement/populations-legales/france-departements.asp?annee=2013>

Dans toutes ces données, ne garder que la colonne qui nous intéresse : Population totale.

On a maintenant une seule colonne donnant la population totale des différentes villes de l'Isère. Pour chacune de ces valeurs, on veut récupérer le premier chiffre (la mantisse). Pour ce faire, on utilisera sans justification le fait que le premier chiffre (aussi appelé mantisse) d'un nombre $x \in \mathbb{R}$ s'exprime en langage Excel :

$$\text{ENT}(10^{(\text{LOG}(x) - \text{ENT}(\text{LOG}(x)))})$$

(ENT désigne la partie entière et LOG le logarithme en base 10).

Calculer la fréquence d'apparition de 1,2,...,9 en tant que premier chiffre. Tracer ces valeurs sur une courbe.

Question 2.1. *La courbe obtenue confirme-t-elle votre conjecture faite à la question 1.1 ?*

3 Comparaison avec la loi de Benford

La loi de Benford est une loi prédisant que le chiffre i apparaît avec une fréquence de $f(i) = \text{LOG}(1 + \frac{1}{i})$.

Question 3.1. *Tracer $f(i)$ en fonction de i . Quel chiffre reviendrait le plus souvent si cette loi est vérifiée ?*

Question 3.2.

Dans le tableur, en supposant cette loi, donner le nombre d'occurrence théorique de chaque premier chiffre (pour la collecte des données de la partie précédente).

Question 3.3. *Représenter sur un même graphique les résultats de la fréquence expérimentale et théorique. Sont-ils proche ?*

Question 3.4. *Reprendre cette étude pour une autre série de valeur.*

Question 3.5. *Supposons que l'on ait la taille en cm des élèves du lycée Pablo Neruda. Quelle est la répartition du premier chiffre ?*

Même question avec les numéros de téléphones de l'annuaire.

Quelle condition peut-on conjecturer pour que la loi de Benford soit effectivement vérifiée ?

4 A finir chez soi

A faire au choix (au moins) l'une des deux questions suivantes. (La seconde est plus dure)

Question 4.1. *Reprendre l'analyse avec une autre série de valeur (issue de l'INSEE ou autre base de donnée en ligne). Retrouve-t-on la loi de Benford ?*

Question 4.2. *Reprendre l'étude précédente avec non pas le premier chiffre mais le dernier chiffre. Montrer que le dernier chiffre est le reste dans la division euclidienne par 10. Retrouve-t-on la loi de Benford ?*