.

Thème 1 : Statistiques

Avant-Propos : Conformément au programme officiel, ce thème du programme sera traité alternativement avec Python ainsi qu'un tableur.

Formules de calcul avec Python

Exercice I : un exemple à traiter

Dans cet exercice, on exploite les données de l' **Exemple 1** du cours :

nombre d'enfants	0	1	2	3	4	5
Effectifs	50	23	10	14	2	1
Effectifs cumulés croissants	50	73	83	97	99	100

1. Identifier le caractère étudié. Est-il quantitatif ?
2. Déclarer une variable matricielle $TabE = \begin{pmatrix} 0 & 1 & 2 & 3 & 4 & 5 \\ 50 & 23 & 10 & 14 & 2 & 1 \end{pmatrix}$ qui vous permettra de traiter les données
3. Compléter la matrice avec les lignes de fréquences et fréquences cumulées croissantes.

nombre d'enfants	0	1	2	3	4	5
Fréquences						
Fréquences cumulées croissantes						

4. Créer la courbe représentative des fréquences cumulées croissantes.

Exercice II : programmes de calculs d'indicateurs

Le but de cet exercice est de retrouver les algorithmes de calcul des indicateurs statistiques classiques
Pour tester vos programmes, on donnera un tableau abstrait de test :

Valeur	2	3	4	5
Effectif	3	14	7	8

Ainsi que le relevé des notes de l'interrogation I_1 :

Note	0	0,5	1	1,5	2	2,5	3	3,5	4	4,5	5,5	6	6,5	7,5
Effectif	2	5	2	2	2	3	3	4	2	2	1	1	2	2

1. Encoder ces deux tableaux sous forme de matrices
2. A partir des lignes (incomplètes) qui suivent, programmer un calcul de la moyennne :

```
def moyenne(T):
        n=size(T)
        n=n[1]
        M=0
        for k in range(n):
                M=M+T[k,0]*T[k,1]
.
.
.
        return(M)
```

puis tester sur les exemples demandés

3. Programmer une fonction `etand` qui calcule l'étendue puis tester sur les exemples.
4. A partir des lignes (incomplètes) qui suivent, programmer un calcul de la variance :

```
def varian(T):
        n=size(T)
        n=n[1]
        V=0
        for k in range(n):
                V=V+(..... -np.mean(T))**2
.
.
.
        return(M)
```

puis tester sur les exemples.

5. Programmer une fonction `ecrtyp` qui calcule l'écart-type à partir de la variance puis tester sur les exemples.
6. Programmer une fonction `coeffvar` qui calcule le coefficient de variation à partir de la variance et de l'écart-type puis tester sur les exemples.

Exercice III : statistiques bivariées

Nous rentrons dans la thématique officielle ECT-2 telle que présentée au programme.
Dans cette partie, on pourra utiliser directement les commandes permettant d'obtenir les indicateurs statistiques univariés vus en ECT-1 de la bibliothèque `numpy` *:*

indicateur :	moyenne	écart type	variance	minimum	maximum	somme	somme cumulée	médiane
commande :	`np.mean`	`np.std`	`np.var`	`np.min`	`np.max`	`np.sum`	`np.cumsum`	`np.median`

Introduction : petite suite de points

On commence par traiter les données de **Exemple 11** du cours :

individu p	Awa	Massy	Julie	Alex	Sandra	Karima	Iliès
jour $x_k =$	12	7	8	30	29	4	6
mois $y_k =$	3	12	6	4	12	2	8

On remarquera que, dans cet exemple, on considère implicitement dans la description que l'effectif de $(x_i; y_j)$ est nul dès que $i \neq j$.

La covariance se résume donc à : $cov(X;Y) = \dfrac{1}{N}\displaystyle\sum_{k=1}^{N}(x_k - \bar{x})(y_k - \bar{y})$ où N est le nombre d'individus du tableau.

1. Encoder les données sous forme de matrice qui sera nommée `ExOnz`
2. Encoder $X = (x_k)$ et $Y = (y_k)$ associés sous forme de listes des valeurs rencontrées (on conservera l'ordre d'énumération).
3. Déterminer les moyennes et écart-type de chaque série de données
4. Mettre en place un programme qui calcule la covariance à l'aide de la définition et d'une boucle `for`.

Tableau croisé à effectifs

Dans cet exercice, le tableau de données utilisé pour tester les commandes sera celui de **Exemple 5** du tableau des effectifs du couple (prix ; masse) d'un panier de biens d'une boutique :

prix / masse :	20 g	100 g	250g	500g	1000g
5 (euros)	23	15	2	0	0
7 (euros)	19	26	12	7	1
8.50 (euros)	13	21	35	22	14
9 (euros)	6	15	24	30	34
9.99 (euros)	1	3	9	24	5

On posera $M = [20; 100; 250; 500; 1000]$ le ligne des masses, la notation m_i désignant alors la $i$ème masse de cette liste et on définit P de façon analogue avec les prix.

1. Encoder ce tableau sous forme d'une matrice notée `Tabex`. Attention ! la ligne 0 contiendra les valeurs de masse, la colonne 0 les prix.
   La valeur stockée en coefficient [0 ;0] sera choisie arbitrairement et ne devra pas être employée dans un calcul
2. A l'aide de Python, déterminer les effectifs marginaux de chaque modalité du caractère *masse*
3. A l'aide de Python, déterminer les effectifs de masse conditionnellement à *8,50 euros* puis les fréquences de masse conditionnelles associées.
4. **vers le calcul de covariance**
   (a) Définir une variable `Pbar` qui stocke la valeur de prix moyen.
   (b) Définir une variable `Mbar` qui stocke la valeur de masse moyenne.
   (c) Produire un nouveau tableau `Covex` tel que `Covex`$[i; j]$ stocke $(m_i - \bar{m})(p_j - \bar{p}) \times \text{eff}(E((m_i; p_j)))$
   (d) Finaliser le calcul de la covariance associée à ce tableau de valeurs.
   (e) Peut-on dire que les prix, dans cette boutique, sont corrélés aux masses des articles vendus ?

Ajustement linéaire par méthode des moindres carrés

Le but de l'exercice est de mettre en place un programme de calcul des paramètres de la droite de regression affine par méthode des moindres carrés.
L'usage concret sur des exemples réels et la lecture des interprétations se fera par le biais d'un tableau dans d'autres séances.

Le tableau qui suit fournit les prix moyens de l'essence à la pompe, observés en fin de semestre chaque année entre 2020 et 2023, en euros par litre (au centime près) en France

Semestre :	S1-2020	S2-2020	S1-2021	S2-2021	S1-2022	S2-2022	S1-2023	S2-2023
Rangs $x_k =$	1	2	3	4	5	6	7	8
Effectifs $y_k =$	1,22	1,22	1,43	1,54	2,09	1,78	1,70	1,77

Source : https ://www.insee.fr/fr/statistiques/serie/000442588

1. Déterminer le point moyen sous la forme d'un vecteur à deux valeurs $G = [$`xbar`$;$ `ybar`$]$
2. Organiser le calcul de covariance de cette série (voir les similitudes avec **exemple 11**)
3. Déterminer m et p tels que présentés dans le cours
4. Déclarer la fonction `RegPompe` définie par $x \mapsto mx + p$
5. Faire s'afficher dans une fenêtre graphique le nuage de points associé au tableau de données, la droite de regression de y en x et le point moyen.
6. Reprendre la démarche en intervertissant les rôles de x et de y pour obtenir la droite de regression de x en y.