

Question de cours (3 points)

Soient $\{x_i\}_{1 \leq i \leq n}$ et $\{y_i\}_{1 \leq i \leq n}$ les deux séries statistiques d'observations des deux variables X et Y respectivement.

Les deux variables X et Y sont liées par la relation : $Y = \frac{X - \bar{x}}{\sigma_X}$.

En utilisant les propriétés de la moyenne et la variance et en détaillant les étapes, calculer $\bar{y}$ et $\text{Var}(Y)$.

Exercice 1 (8 points) (Donner les résultats avec 3 chiffres après la virgule)

Dans une enquête sociale publiée par l'INSEE en 2016 sur la satisfaction de la population concernant les relations avec les amis, on cherche à étudier la liaison entre l'âge en années (variable notée X) et une note de satisfaction sur 10 (variable notée Y).

Sur un échantillon de 7 individus interrogés, on donne les résumés numériques suivants :

$$\sum_{i=1}^7 x_i = 364 \quad \sum_{i=1}^7 y_i = 56,5 \quad \sum_{i=1}^7 x_i^2 = 21\,728 \quad \sum_{i=1}^7 y_i^2 = 456,47 \quad \sum_{i=1}^7 x_i y_i = 2\,908$$

- 1) Donner le type de chaque variable.
- 2) Calculer la moyenne et la variance de la variable X .
- 3) Calculer la moyenne et la variance de la variable Y .
- 4) Calculer la covariance entre les deux variables.
- 5) En déduire la valeur du coefficient de corrélation linéaire entre les deux variables.
- 6) Déterminer les valeurs des coefficients estimés $\hat{a}$ et $\hat{b}$ de la droite de régression $y = \hat{a}x + \hat{b}$.
- 7) Prédire la note de satisfaction d'un individu âgé de 40 ans.

Exercice 2 (9 points) (Donner les résultats avec 2 chiffres après la virgule)

Le tableau ci-dessous donne la répartition de 1000 individus selon leur diplôme et leur utilisation d'internet (extrait d'une enquête de l'INSEE sur l'utilisation d'internet en France en 2017).

	pas d'internet	rarement	tous les jours	total
Aucun diplôme ou CEP	100	35	65	200
BEP, CAP ou brevet des collèges	18	20	62	100
Baccalauréat	25	78	397	500
Diplôme supérieur au baccalauréat	8	19	173	200
total	151	152	697	1000

Pour la cellule (i, j) du tableau, on note n_{ij} l'effectif observé, t_{ij} l'effectif théorique attendu en cas d'indépendance parfaite, e_{ij} l'écart entre les effectifs observé et attendu.

- 1) Comment appelle-t-on ce tableau ?
- 2) Donner le tableau des profils-colonnes et conclure sur l'existence ou non d'une liaison entre les deux variables (justifier votre réponse).
- 3) On note $f_{i\cdot}$ et $f_{\cdot j}$ les fréquences marginales des deux variables (notations du cours).
 - a) Calculer les fréquences marginales $f_{i\cdot}$ et interpréter $f_{3\cdot}$.
 - b) Calculer les fréquences marginales $f_{\cdot j}$ et interpréter $f_{\cdot 2}$.
- 4) Calculer l'écart entre les effectifs observé et attendu pour les couples (pas d'internet, aucun diplôme ou CEP) et (rarement, Diplôme supérieur au baccalauréat). Commenter le sens de l'association pour chacun de ces couples.
- 5) On donne la valeur $\chi^2 = 273,93$.
 Calculer le coefficient de Cramer et qualifier l'intensité de la liaison entre les deux variables.