

Questions de cours :

Soient $x_1, x_2, \dots, x_n, y_1, \dots, y_n$ des observations (c'est-à-dire des nombres réels). On note $x = (x_1, \dots, x_n)^\top$ et $y = (y_1, \dots, y_n)^\top$ les vecteurs des observations.

1. Donner la définition de la variance de x .
2. Donner la définition de l'écart-type de x .
3. Donner une formule de la covariance entre x et y .

Exercice 1 :

Suite à l'appel du Noyau Populaire Autonome et du Comité de Grève Turbulente, la France se prépare à une grève générale. Le tableau (fictif) suivant donne les effectifs de la population française soutenant ou ne soutenant pas la grève générale, en fonction de l'âge.

Tranche d'âge	Soutient la grève (Pro)	Ne soutient pas la grève (Anti)	Total
(1) 15-24	5 500 000	1 700 000	7 200 000
(2) 25-34	4 700 000	2 000 000	6 700 000
(3) 35-44	4 000 000	2 700 000	6 700 000
(4) 45-54	3 300 000	4 000 000	7 300 000
(5) 55-64	2 400 000	4 200 000	6 600 000
Total	19 900 000	14 600 000	34 500 000

1. Depuis le tableau de contingence, calculer les valeurs suivantes (*On pourra laisser le résultat sous la forme d'une fraction*) :
 - la probabilité pour qu'une personne soit jeune (15-24 ans) et soutienne la grève,
 - la probabilité pour qu'une personne ayant plus de 45 ans ne soutienne pas la grève,
 - la probabilité pour qu'une personne soutenant la grève ait entre 25 et 34 ans.
2. Dans cette question, on note x_1 la moyenne d'âge en années dans la tranche 15-24 ans, x_2 la moyenne d'âge dans la tranche 25-34, x_3 la moyenne pour les 35-44, x_4 la moyenne pour les 45-54 et x_5 la moyenne pour les 55-64. On note $n_{P,i}$ le nombre de personnes dans la catégorie (i) qui soutiennent la grève (**Pro**) et $n_{A,i}$ le nombre de personnes dans la catégorie (i) qui ne soutiennent pas la grève (**Anti**).
 Donner une formule permettant de calculer $\bar{x}_P$ la moyenne d'âge des personnes soutenant la grève, et une formule permettant de calculer $\bar{x}_A$ la moyenne d'âge des personnes ne soutenant pas la grève. A-t-on assez d'informations pour calculer ces moyennes ? (Justifier.)
3. Dire et **justifier** si les affirmations suivantes sont correctes :
 - Plus on est jeune, plus on a de chances de soutenir la grève.
 - Si on ne prend pas en compte les moins de 24 ans, la majorité des personnes ne soutiennent pas la grève.

Exercice 2 :

Soient $x_1, \dots, x_n$ des nombres réels. Montrer que la moyenne

$$\bar{x} = \frac{1}{n} \sum_{k=1}^n x_k$$

minimise l'écart quadratique défini par, pour tout réel $u \in \mathbb{R}$,

$$C_2(u) = \sum_{k=1}^n (x_k - u)^2.$$

Correction

Questions de cours :

On note $\bar{x} := \frac{1}{n} \sum_{i=1}^n x_i$ la moyenne de x et $\bar{y} := \frac{1}{n} \sum_{i=1}^n y_i$ la moyenne de y .

1. $\text{var}(x) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$
2. $\sigma(x) = \sqrt{\text{var}(x)}$
3. $\text{cov}(x, y) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$

Exercice 1 :

1. — La probabilité pour qu'une personne soit jeune (15-24 ans) et soutienne la grève est $\frac{55}{345}$.
— La probabilité pour qu'une personne ayant plus de 45 ans ne soutienne pas la grève est $\frac{33+24}{73+66} = \frac{57}{139}$.
— La probabilité pour qu'une personne soutenant la grève ait entre 25 et 34 ans est $\frac{47}{199}$.

2.

$$\bar{x}_P = \frac{1}{\sum_{j=1}^5 n_{P,j}} \sum_{i=1}^5 n_{P,i} x_i, \quad \bar{x}_A = \frac{1}{\sum_{j=1}^5 n_{A,j}} \sum_{i=1}^5 n_{A,i} x_i$$

On ne dispose pas des moyennes d'âge dans chaque groupe (les x_i), donc on ne peut pas calculer exactement les moyennes $\bar{x}_P$ et $\bar{x}_A$. (En revanche, comme l'ont fait remarquer plusieurs personnes, on peut obtenir un résultat approché en estimant que les moyennes dans les groupes seraient proche de la moyenne entre les âges extrêmes de la catégorie, par exemple environ 20 ans pour la catégorie 15-24.)

3. — Plus on est jeune, plus on a de chances de soutenir la grève : **Vrai**, parmi les proportions de personnes favorables à la grève dans deux classes d'âges, la plus élevée est toujours celle de la classe la plus jeune.
— Si on ne prend pas en compte les moins de 24 ans, la majorité des personnes ne soutiennent pas la grève : **Faux**, en enlevant les moins de 24 ans, il y a $19,9-5,5=14,4$ millions de personnes pro-grève, contre $14,6-1,7=12,9$ millions de personnes anti-grève.

Exercice 2 :

La fonction C_2 est polynomiale donc infiniment dérivable. Si elle admet un minimum en un point u_0 , alors sa dérivée s'annule en ce point (**MAIS** a priori, ce n'est pas parce que sa dérivée s'annule en un point que c'est automatiquement un minimum). Pour tout réel u , on a

$$\begin{aligned} C'_2(u) = 0 &\iff \sum_{k=1}^n 2(u - x_k) = 0 \\ &\iff nu - \sum_{k=1}^n x_k = 0 \\ &\iff u = \bar{x} \end{aligned}$$

De plus, pour tout réel u , $C''_2(u) = 2n > 0$. La fonction C_2 est donc convexe, et $\bar{x}$ est donc la valeur pour laquelle cette fonction atteint son minimum.

Autre possibilité de rédaction :

Pour tout réel u ,

$$\begin{aligned} C_2(u) &= \sum_{k=1}^n (x_k - \bar{x} + \bar{x} - u)^2 \\ &= \sum_{k=1}^n (x_k - \bar{x})^2 + 2(\bar{x} - u) \underbrace{\sum_{k=1}^n (x_k - \bar{x})}_{=0} + n(\bar{x} - u)^2 \\ &= \sum_{k=1}^n (x_k - \bar{x})^2 + n(\bar{x} - u)^2 \end{aligned}$$

Cette quantité est une somme de termes positifs, elle est donc minimale lorsque le terme de droite s'annule, c'est-à-dire lorsque $u = \bar{x}$.