

CORRECTION CC DE MI-SEMESTRE DE STATISTIQUES

Selon l'Organisation Mondiale de la Santé, il y a surpoids quand l'Indice de Masse Corporelle (IMC) est égal ou supérieur à 25 et il y a obésité quand l'IMC est égal ou supérieur à 30.

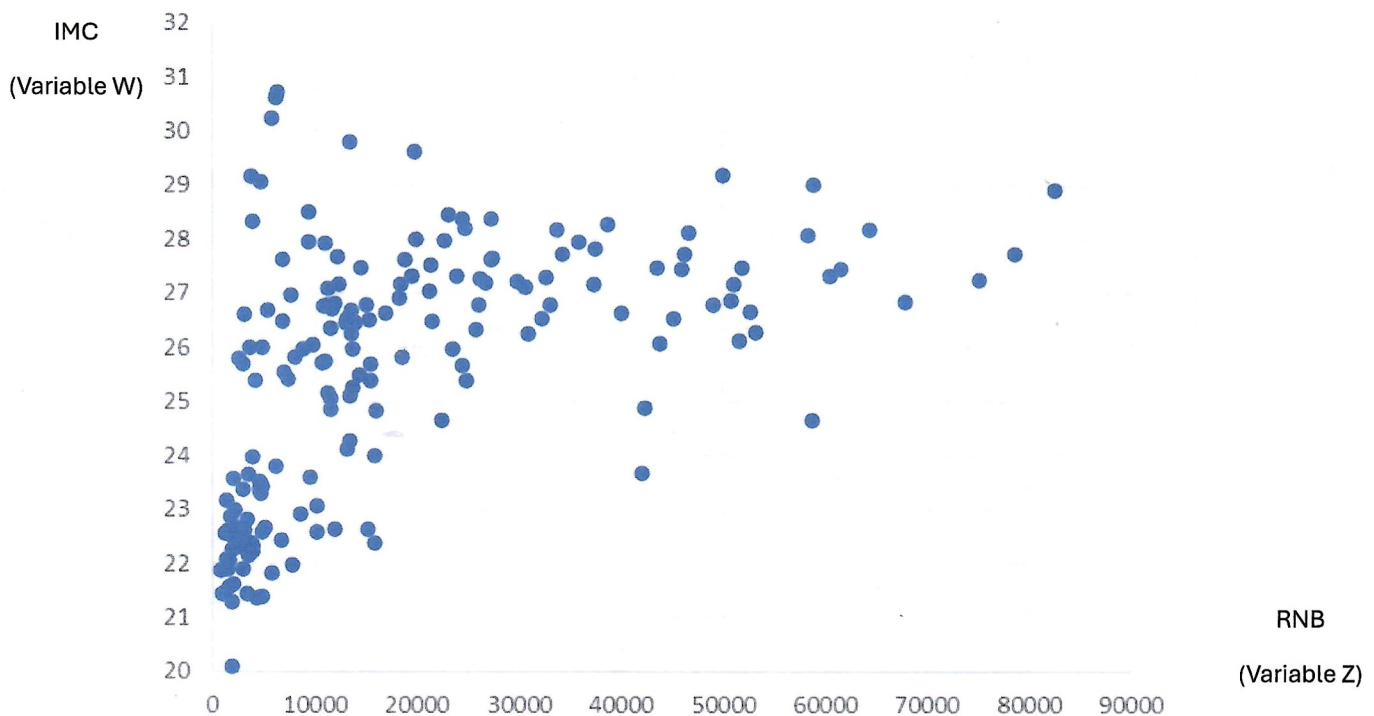
$$IMC = \frac{\text{poids (en kg)}}{\text{taille}^2 \text{ (en m)}}$$

On dispose de l'IMC moyen des hommes (sans unité, en ordonnée, notée W) et du revenu national brut par habitant (en dollars, en abscisses, notée Z) de 176 pays (données de la Banque Mondiale, 2016).

IMC (en kg/m²) Classification OMS

Moins de 16,5	Dénutrition
Entre 16,5 et 18,5	Maigre
Entre 18,5 et 25	Valeur de référence
Entre 25 et 30	Surpoids
Entre 30 et 35	Obésité modérée
Entre 35 et 40	Obésité sévère
Au-delà de 40	Obésité massive

Z / W	[18,5 ; 24,9]	[24,9 ; 29,9]	[29,9 ; 40]	TOTAL
[0 ; 30 000)	60	75	3	138
[30 000 ; 60 000)	3	28	0	31
[60 000 ; 90 000)	0	7	0	7
TOTAL	63	110	3	176



Indice de masse corporelle (IMC) moyen chez les hommes en fonction du RNB par habitant en dollars, dans 176 pays du monde en 2016

Nous disposons des données suivantes :

$$\sum_{i=1}^N w_i z_i = 87\,948\,773,60$$

$$\bar{w} = 25,55$$

$$\bar{z} = 18\,682,56$$

$$s_w = 2,55$$

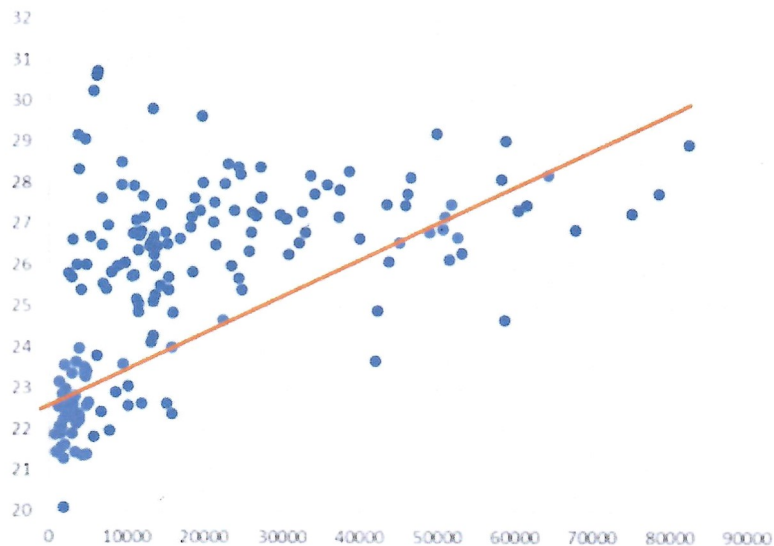
$$s_z = 18\,267,42$$

Donner une première interprétation du nuage de points.

Ce nuage représente l'IMC moyen des hommes d'un pays donné (en ordonnée) en fonction du niveau de richesse de ce pays, approché par le RNB/hab. (en abscisse).

De façon générale, quelque soit le niveau de richesse du pays, les IMC sont normaux/standards voire élevés dans les 176 pays de notre échantillon puisque seuls 63 pays présentent un IMC moyen des hommes inférieur à 25.

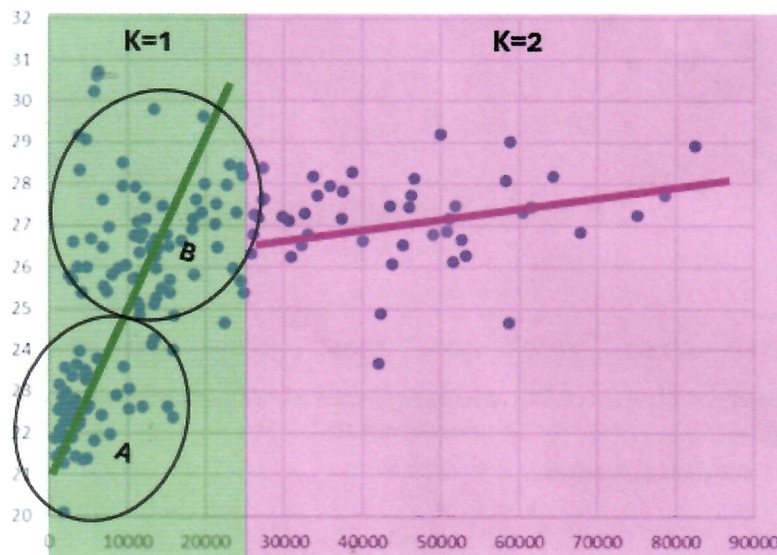
Dans notre échantillon, il semble y avoir une relation positive entre IMC et RNB/hab : plus le RNB est élevé, plus l'IMC semble élevé. Cela permettrait de poser l'hypothèse que le surpoids et l'obésité seraient plus important dans les pays riches. Cette relation peut être synthétisée avec la droite ci-dessous.



Toutefois, notre échantillon peut être divisé en 2 groupes :

$K = 1$ avec les pays dont le RNB/hab est inférieur à 25K. Dans ce cas, il semble exister une relation linéaire avec une forte pente (IMC qui croît rapidement quand RNB/hab augmente).

$K = 2$ avec les pays dont le RNB/hab est supérieur à 25K. Dans ce cas, il semble aussi exister un plateau : quel que soit l'augmentation du RNB/hab, l'IMC stagne autour de 27 [surpoids].



Enfin, si on zoom plus en détails sur le groupe de pays avec un RNB/hab inférieur à 25K, on distingue 2 sous-groupes : un sous-groupe A avec un IMC moyen normal et un sous-groupe B avec un IMC moyen en surpoids.

Sur l'ensemble de nos données, on s'attend donc à avoir un coefficient de corrélation linéaire plutôt modéré, voire faible et positif entre W et Z .

Calculer la covariance empirique entre W et Z.

$$s_{wz} = \frac{1}{N-1} \sum_{i=1}^N w_i z_i - \frac{N}{N-1} \bar{w} \bar{z} = \frac{1}{175} \times 87\,948\,773,60 - \frac{176}{175} \times 25,55 \times 18\,682,56 \\ = 502\,564,4206 - 480\,067,0618 = 22\,497,36$$

Donc

$$s_{wz} = 22\,497,36$$

La covariance entre l'IMC moyen des hommes et le RNB/hab est de 22 497,36 [kg.m⁻².\$].

Calculer le coefficient de corrélation linéaire entre W et Z. Interpréter au seuil de 0,3.

$$r_{wz} = \frac{s_{wz}}{s_w \times s_z} = \frac{22\,497,36}{2,55 \times 18\,267,42} = 0,48$$

Le coefficient de corrélation linéaire est supérieur en valeur absolue au seuil de 0,3. On conclue donc qu'il existe une relation linéaire entre l'IMC et le RNB/hab. Toutefois, il est plus proche du seuil que de 1 (en valeur absolue), cette relation est donc assez faible/modérée.

Le coefficient de corrélation linéaire est positif : on conclue donc qu'il existe une relation linéaire positive entre nos variables : plus le RNB/hab du pays est élevé, plus l'IMC moyen des hommes de ce pays est élevé.

Calculer et interpréter les estimateurs des paramètres du modèle suivant par les MCO : $w_i = a + bz_i + u_i$

On cherche les estimateurs du modèle suivant : $\hat{w}_i = \hat{a} + \hat{b}z_i$

$$\hat{b} = \frac{s_{wz}}{s_z^2} = \frac{22\,497,36}{18\,267,42^2} = 6,74 \times 10^{-5}$$

Une augmentation d'un dollar de RNB/hab mène à une augmentation de $6,74 \times 10^{-5}$ de l'IMC moyen des hommes.

$$\hat{a} = \bar{w} - \hat{b}\bar{z} = 25,55 - 6,74 \times 10^{-5} \times 18\,682,56 = 24,29$$

Selon ce modèle, l'IMC d'un pays avec un RNB/hab égal à zéro est de 24,29 en moyenne.

Le modèle final d'ajustement linéaire est donc

$$\hat{w}_i = 24,29 + 6,74 \times 10^{-5} z_i$$

Tracer la droite relative à ce modèle sur le graphique de l'énoncé.

Pour $Z = 0$

$$W = 24,29 + 6,74 \times 10^{-5} \times 0 = 24,29$$

Soit A (0 ; 24,29)

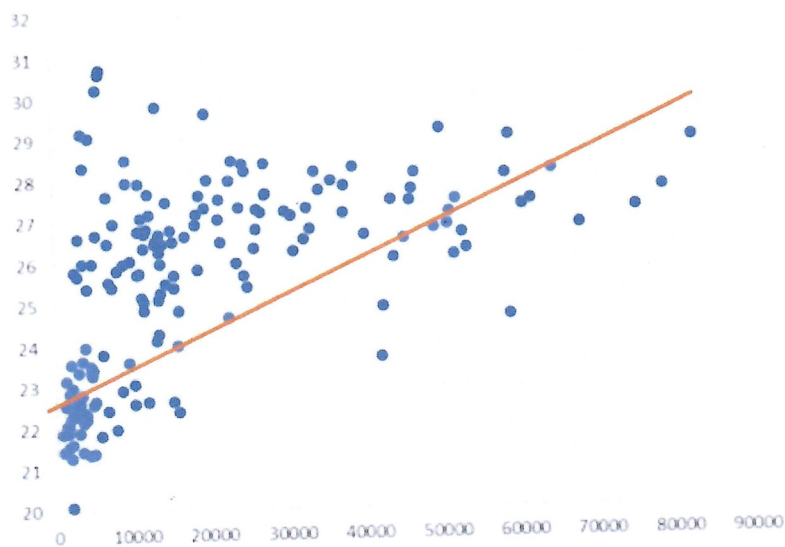
Pour $Z = 90\,000$

$$W = 24,29 + 6,74 \times 10^{-5} \times 90\,000 = 30,36$$

Soit B (90 000 ; 30,36)

De façon générale, quelque soit le niveau de richesse du pays, les IMC sont normaux/standards voire élevés dans les 176 pays de notre échantillon puisque seuls 63 pays présentent un IMC moyen des hommes inférieur à 25.

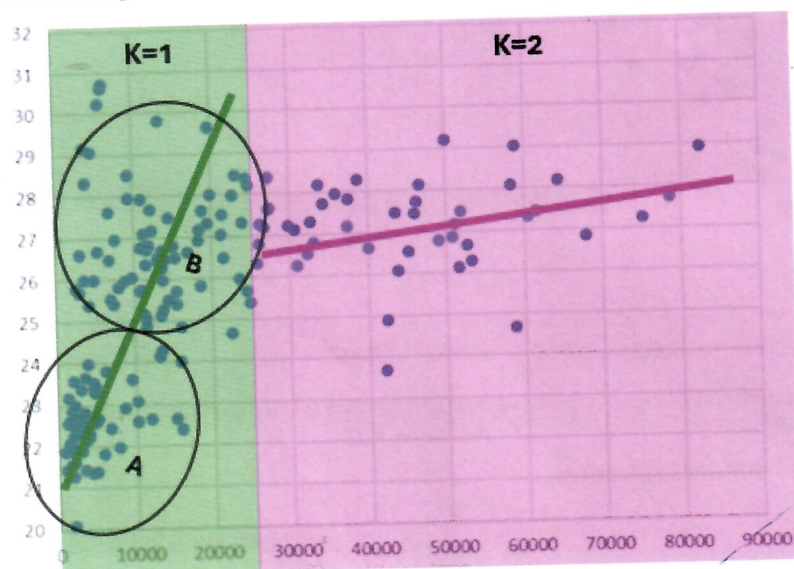
Dans notre échantillon, il semble y avoir une relation positive entre IMC et RNB/hab : plus le RNB est élevé, plus l'IMC semble élevé. Cela permettrait de poser l'hypothèse que le surpoids et l'obésité seraient plus importants dans les pays riches. Cette relation peut être synthétisée avec la droite ci-dessous.



Toutefois, notre échantillon peut être divisé en 2 groupes :

$K = 1$ avec les pays dont le RNB/hab est inférieur à 25K. Dans ce cas, il semble exister une relation linéaire avec une forte pente (IMC qui croît rapidement quand RNB/hab augmente).

$K = 2$ avec les pays dont le RNB/hab est supérieur à 25K. Dans ce cas, il semble aussi exister un plateau : quel que soit l'augmentation du RNB/hab, l'IMC stagne autour de 27 [surpoids].



Enfin, si on zoom plus en détails sur le groupe de pays avec un RNB/hab inférieur à 25K, on distingue 2 sous-groupes : un sous-groupe A avec un IMC moyen normal et un sous-groupe B avec un IMC moyen en surpoids.

Sur l'ensemble de nos données, on s'attend donc à avoir un coefficient de corrélation linéaire plutôt modéré, voire faible et positif entre W et Z.