Classe: Tle IH

Période: du 16 Avril 2020 au 23 Avril 2020.

Thème: STATISTIQUES

Leçon 2: SERIES STATISTIQUES A DEUX CARACTERES.

Activité:

On a relevé le poids en kg et la taille en cm de 10 élèves de la Tle IH. Les résultats sont les suivants:

Poids (x_i)	65	68	62	62	68	62	65	68	71	74
Taille (y_j)	165	168	174	168	171	174	174	171	171	174

2-1 Organisation des données:

Les données de notre activité permettent de définir deux séries statistiques à deux caractères qui sont:

Poids (x_i)	62	65	68	71	74
Effectif (n_i)	3	2	3	1	1

Taille (y_j)	165	168	171	174
Effectif (n_j)	1	2	3	4

Ces deux séries sont appelées séries marginales de la série double.

Posons $X = \{62; 65; 68; 71; 74\}$

et $Y = \{165; 168; 171; 174\}$

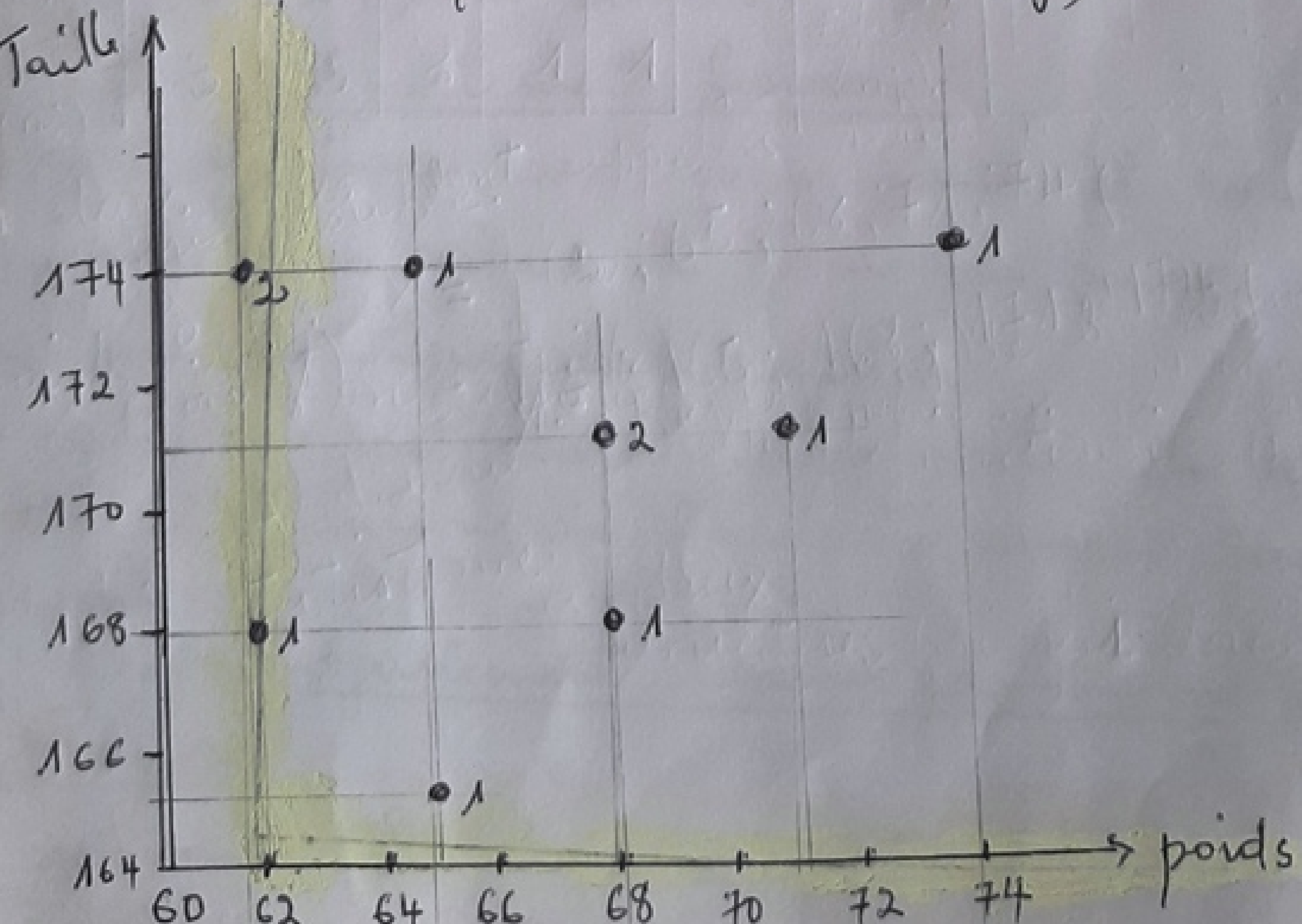
A chaque couple $(x; y)$ de l'ensemble $X \times Y$ (X croix Y), on associe le nombre x d'élèves de poids x et de taille y. On définit ainsi une série statistique à deux caractères (ou série double).

On peut aussi représenter cette série double dans un tableau à double entrée :

Taille \ Poids	62	65	68	71	74	Total
165	0	1	0	0	0	1
168	1	0	1	0	0	2
171	0	0	2	1	0	3
174	2	1	0	0	1	4
Total	3	2	3	1	1	10

2-1-1 Nuage de points :

Le plan est muni d'un repère orthogonal.
Représentons chaque modalité $(x;y)$.



Ce graphique est appelé nuage de points.

2-1-2 Points pondérés :

Par convention, on indiquera à droite de chaque point du nuage, l'effectif de son couple de coordonnées. L'ensemble des points obtenus est appelé ensemble de points pondérés.

2-1-3 Points moyens d'un nuage de points :

Dans notre activité,
le poids moyen des élèves est :

$$m_1 = \frac{62\times 3 + 65\times 2 + 68\times 3 + 71\times 1 + 74\times 1}{10} = \frac{665}{10} = 66,5$$

La taille moyenne des élèves est :

$$m_2 = \frac{165\times 1 + 168\times 2 + 171\times 3 + 174\times 4}{10} = \frac{1710}{10} = 171$$

Le point $G\begin{pmatrix}66,5\\171\end{pmatrix}$ est appelé point moyen du nuage à la série double.

2-2. Ajustement linéaire :

Il s'agit de tracer une courbe simple et régulière qui épouse au mieux la forme du nuage associé à la série statistique double étudiée.
Pour cela, on peut choisir deux points du nuage et écrire l'équation de la droite passant par ces deux points.

2-2-1 Ajustement linéaire par la méthode de Mayer :

Pour déterminer la droite d'ajustement d'un nuage par la méthode de Mayer, on peut :

- Partager le nuage de points en deux groupes de même importance (à une unité près si le nombre de points est impair) dans l'ordre où les points se présentent.
- Calculer les coordonnées des points moyens G_1 et G_2 de chaque groupe de points.
- Écrire une équation cartésienne de la droite passant par G_1 et G_2. $(G_1;G_2)$ est la droite de Mayer ajustant le nuage de points de la série double étudiée.

2-2-2 <u>Ajustement affine par la méthode des moindres carrés:</u>

2-2-2-1 <u>Définition:</u>

$\mathcal{P}$ est une population d'effectif N.
X et Y deux caractères étudiés sur la population $\mathcal{P}$.
$\{x_1; x_2; \ldots; x_i\}$ l'ensemble des modalités de X.
$\{y_1; y_2; \ldots; y_j\}$ l'ensemble des modalités de Y.
$\bar{x}$ est la moyenne de la série du caractère X.
$\bar{y}$ est la moyenne de la série du caractère Y.
n_{ij} l'effectif du couple de modalité $(x_i; y_j)$
On appelle covariance de la série statistique double de caractère $(X;Y)$, le nombre réel noté

$$\operatorname{cov}(X;Y) = \frac{1}{N}\sum_{i=1}^{n} n_{ij}\, x_i y_j - \bar{x}\bar{y}$$

$$= \frac{1}{N}\sum n_{ij}(x_i - \bar{x})(y_j - \bar{y}).$$

Propriété:

P est une population d'effectif N.
X et Y deux caractères étudiés sur la population P.
$\{x_1; x_2; \dots; x_i\}$ l'ensemble des modalités de X.
$\{y_1; y_2; \dots; y_j\}$ l'ensemble des modalités de Y.
$\overline{x}$ la moyenne de la série de caractère X.
$\overline{y}$ est la moyenne de la série de caractère Y.
n_{ij} l'effectif du couple de modalités $(x_i; y_j)$.
$V(X)$ la variance de la série de caractère X
$V(Y)$ la variance de la série de caractère Y.
$\text{cov}(X;Y)$ la covariance de la série double de caractères X et Y.

La droite de regression $D_{x/y}$ de X en Y a pour équation $x-\overline{x} = a'(y-\overline{y})$ avec $a' = \dfrac{\text{Cov}(X;Y)}{V(Y)}$

La droite de regression $D_{y/x}$ de Y en X a pour équation $y-\overline{y} = a(x-\overline{x})$ avec $a = \dfrac{\text{Cov}(X;Y)}{V(X)}$ encore appelé la droite des moindres carrés. Cette droite passe par $G\begin{pmatrix}\overline{x}\\ \overline{y}\end{pmatrix}$ point moyen du nuage.

On appelle coefficient de corrélation linéaire des séries X et Y; le nombre réel noté r défini par

$$r = \frac{\text{Cov}(X;Y)}{\sqrt{V(X)\cdot V(Y)}} = \sqrt{aa'}$$

TAF: n° 15 p 128 Majors en maths Tle D.