**Université Mohamed Boudiaf - M'sila**

**Faculté des Sciences et Technologies**

**Départements de Hydraulique**

**Licence (2/HYDROLIQUE)**

**Année Universitaire 2020/2021**

# Module : ( Probabilités et Statistiques )

**U.E Methodologique; Crédits: 04; Coefficient: 02**

**Contrôle continue 40%; Examen: 60%**

# Partie A: Statistiques

# Chapitre 1

# Définitions de base-Séries statistiques à une variable

# Chapitre 2

# Séries statistiques à deux variables

**Définition 2.0.1** *On appelle série statistique à deux varible $X$ et $Y$ toute liste de couples $(x_1, y_1)\,;(x_2, y_2)\,;...;(x_n, y_n)$. $n$ est le nombre de ces couples.*

| $x_i$ | $x_1$ | ... | $x_n$ |
|---|---|---|---|
| $y_i$ | $y_1$ | ... | $y_n$ |

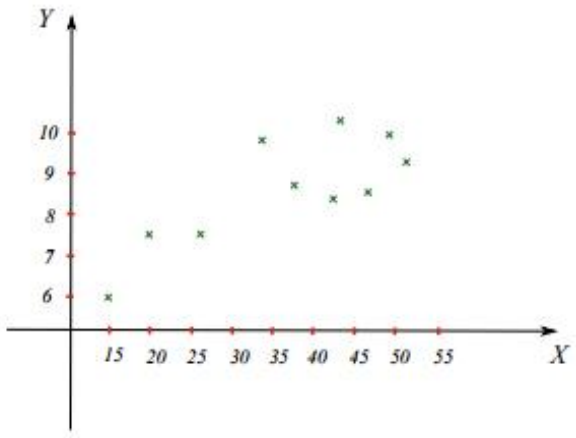
**Exemple 2.0.1** *Nous considérons 10 salariés qui sont observés à l'aide de deux variables (âge) est (salaire), les informations brutes sont données dans le tableau suivant;*

| *Salaire* | *6000* | *7400* | *7500* | *8200* | *8207* | *8900* | *9100* | *9900* | *9950* | *10750* |
|---|---|---|---|---|---|---|---|---|---|---|
| *Age* | *15* | *26* | *20* | *43* | *47* | *37* | *52* | *34* | *50* | *44* |

## 2.1 Nuage de Points:

**Définition 2.1.1** *Etant donné une série statistique, on appelle nuage de points associé l"ensemble des $n$ points $M_1, M_2, ..., M_n$ du plan de coordonnées $(x_1, y_1)\,;(x_2, y_2)\,;...;(x_n, y_n)$.*

**Exemple 2.1.1 (suite)** *Le nuage de points est tracé, à partir des données brutes, dans la figure suivante.*




**Définition 2.1.2** *On appelle point moyenne d'un nuage de points le point G de coordonnées* $(\overline{x}, \overline{y})$ *où* $\overline{x}$ *est la moyenne de* $x_1, x_2, ..., x_n$ *et* $\overline{y}$ *est la moyenne de* $y_1, y_2, ..., y_n$.

$G(\overline{x}, \overline{y})$.

**Exemple 2.1.2 (suite)** $\overline{x} = 7595.7, \overline{y} = 36.8$.

## 2.2 Distribution statistique d'un couple de variables

On appelle distribution statistique du couple $(X, Y)$ le regroupement $\{(x_i, y_i), n_{ij}\}$ où:

$$\begin{cases} x_1, x_2, ..., x_k \text{ les } k \text{ modalités ou valeurs de } X \\ x_1, x_2, ..., x_l \text{ les } l \text{ modalités ou valeurs de } Y \\ n_{ij} \text{ est l'effectif croissé de } (x_i, y_i) \end{cases}$$

### 2.2.1 Tableau de contingence

Représentation de la distribution jointe du couple $(X, Y)$; on utilise un tableau à double entrée appelé **Tableau de contingence.**

| $x/y$ | $y_1$ | ... | $y_j$ | ... | $y_l$ | totaux |
|---|---|---|---|---|---|---|
| $x_1$ | $n_{11}$ | | $n_{1j}$ | | $n_{1l}$ | $n_{1\bullet}$ |
| ... | | | | | ... | ... |
| $x_i$ | $n_{i1}$ | | $n_{ij}$ | | $n_{il}$ | $n_{i\bullet}$ |
| ... | ... | | ... | | ... | ... |
| $x_k$ | $n_{k1}$ | | $n_{kj}$ | | $n_{kl}$ | $n_{k\bullet}$ |
| totaux | $n_{\bullet 1}$ | | $n_{\bullet j}$ | | $n_{\bullet l}$ | $N$ |

$\Longleftarrow$ distribution marginale de la variable $X$ $(x_i; n_{i\bullet})$

$\Uparrow$

distribution marginale de la variable $Y$ $(y_j; n_{\bullet j})$

### 2.2.2 Distribution marginales

On ajoute au tableau de contingence les totaux en ligne et en colonne.

**Définition 2.2.1** *Les $k$ couples $(x_i; n_{i\bullet})$ définissent la distribution marginale de la variable $X$.*

*Les $k$ couples $(y_j; n_{\bullet j})$ définissent la distribution marginale de la variable $Y$*

**Remarque 2.2.1**

$$\begin{cases} n_{i\bullet} = \sum_{j=1}^{l} n_{ij} = \textit{totale de la ligne } i \\ \sum_{i=1}^{k} n_{i\bullet} = N \\ n_{\bullet j} = \sum_{i=1}^{k} n_{ij} = \textit{totale de la ligne } j \\ \sum_{j=1}^{l} n_{\bullet j} = N \end{cases}$$

**Exemple 2.2.1 (suite)** *1. Déterminer le tableau de contingence ($X$ : âge, $Y$ : salaire). Pour l'âge et pour le salaire, former respectivement des classes de pas de 10 ans et de 1000 Da.*

2. *Déterminer l'effectifs marginaux de $X$ et deY.*

*pour le salaire:*

$$\begin{aligned} \textit{Nombre de classe pour le salaire} &= \frac{e}{a_{sal}} = \frac{10750-6000}{1000} = 4.75 \simeq 5 \textit{ classes} \\ \textit{Nombre de classe pour l'âge} &= \frac{e}{a_{âge}} = \frac{52-15}{10} = 3.7 \simeq 4 \textit{ classes} \end{aligned}$$

*Donc*

| Age x/Salaire y | $[6,7[$ | $[7,8[$ | $[8,9[$ | $[9,10[$ | $[10,11[$ | $n_{i\bullet}$ | $f_{i\bullet}$ |
|---|---|---|---|---|---|---|---|
| $[15,25[$ | 1 | 1 | 0 | 0 | 0 | 2 | 0.2 |
| $[25,35[$ | 0 | 1 | 0 | 1 | 0 | 2 | 0.2 |
| $[35,45[$ | 0 | 0 | 2 | 0 | 1 | 3 | 0.3 |
| $[45,55[$ | 0 | 0 | 1 | 2 | 0 | 3 | 0.3 |
| $n_{\bullet j}$ | 1 | 2 | 3 | 3 | 1 | 10 | 1 |
| $f_{\bullet j}$ | 0.1 | 0.2 | 0.3 | 0.3 | 0.1 | 1 | |

**Remarque 2.2.2** *-La fréquence du couple $(x_i, y_i)$ : $f_{ij} = \frac{n_{ij}}{N}$,*

*-La fréquence marginale de $y_j$: $f_{\bullet j} = \frac{n_{\bullet j}}{N} = \sum_{i=1}^{k} f_{ij}$,*

*-La fréquence marginale de $x_i$: $f_{i\bullet} = \frac{n_{i\bullet}}{N} = \sum_{j=1}^{l} f_{ij}$,*

**Exemple 2.2.2 (suite)**

4. Calculer $f_{21}, f_{12}, f_{45}, f_{33}$.

$f_{21} = \frac{n_{21}}{10} = 0 \qquad f_{45} = \frac{n_{45}}{10} = 0$

$f_{12} = \frac{n_{12}}{19} = 0.1 \quad f_{33} = \frac{n_{33}}{10} = 0.2$

5. Déterminer le tableau statistique des deux série marginales $X$ et $Y$.

| $X$ | $[15,25[$ | $[25,35[$ | $[35,45[$ | $[45,55[$ |
|---|---|---|---|---|
| $n_{i\bullet}$ | 2 | 2 | 3 | 3 |
| $f_{i\bullet}$ | 0.2 | 0.2 | 0.3 | 0.3 |
| $x_i$ le centre | 20 | 30 | 40 | 50 |

| $Y$ | $[6,7[$ | $[7,8[$ | $[8,9[$ | $[9,10[$ | $[10,11[$ |
|---|---|---|---|---|---|
| $n_{i\bullet}$ | 1 | 2 | 3 | 3 | 1 |
| $f_{i\bullet}$ | 0.1 | 0.2 | 0.3 | 0.3 | 0.1 |
| $y_i$ le centre | 6.5 | 7.5 | 8.5 | 9.5 | 10.5 |

### 2.2.3 Distributions Conditionelles

**Définition 2.2.2** *On appelle distribution Conditionnelle de $Y$ sachant $X = x_i$:; la distribution $\{(y_1, n_{i1}), ..., (y_l, n_{il})\}$*

*$\Longrightarrow$il s'agit donc de la distribution à une variable donnée par le ième ligne du tableau de contingence (noté $Y/_{X=x_i}$).*

*On peut de même définir la distribution conditionnelle de $X$ sachant $Y = y_j$ : $\{(x_1, n_{1j}), ..., (x_k, n_{kj})\}$*

*$\Longrightarrow$c'est la j-ème colonne du tableau (notée $X/_{Y=y_j}$).*

**Exemple 2.2.3 (suite)** *6. $X/_{Y=[7,8[}$; $Y/_{X=[45,55[}$*

| $X/_{Y=[7,8[}$ | $[15,25[$ | $[25,35[$ | $[35,45[$ | $[45,55[$ |
|---|---|---|---|---|
| $n_i$ | 1 | 1 | 0 | 0 |

| $Y/_{X=[45,55[}$ | $[6,7[$ | $[7,8[$ | $[8,9[$ | $[9,10[$ | $[10,11[$ |
|---|---|---|---|---|---|
| $n_i$ | 0 | 0 | 1 | 2 | 0 |

## 2.3 Covariance

**Définition 2.3.1** *On appelle covariance de la série statistique double de variables $x$ et $y$ le nombre réel*

$$COV(X,Y) = \sigma_{xy} = \frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x})(y_i - \overline{y})$$

*Pour les calculs, on pourra aussi utiliser :*

$$\sigma_{xy} = \frac{1}{N}\sum_{i=1}^{N} x_i y_i - \overline{xy}$$

**Preuve.** (exercice) ■

**Remarque 2.3.1** *1.* $COV(X,X) = Var(X)$.

*2.* $COV(Y,Y) = Var(Y)$.

**Preuve.**

1. $COV(X,X) = \frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x})(x_i - \overline{x}) = \frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x})^2 = Var(X)$.

2. $COV(Y,Y) = \frac{1}{N}\sum_{i=1}^{N}(y_i - \overline{y})(y_i - \overline{y}) = \frac{1}{N}\sum_{i=1}^{N}(y_i - \overline{y})^2 = Var(Y)$.

■

**Exemple 2.3.1 (suite)** *7. Calculer* $COV(X,Y), COV(X,X)$.

$$\begin{aligned} COV(X,Y) &= \frac{1}{N}\sum_{i=1}^{N} x_i y_i - \overline{xy} = \frac{1}{10}(600 \times 15 + ... + 10750 \times 44) - (7595.7 \times 36.8) = 10763.64 \\ COV(X,X) &= \frac{1}{N}\sum_{i=1}^{N} x_i^2 - \overline{x}^2 = \frac{1}{10}\left(15^2 + ... + 44^2\right) - (36.8)^2 = 148.16 \end{aligned}$$

## 2.4 Ajustement linéaire

### 2.4.1 Le problème de l'ajustement affine

Le nuage de points associé à une série statistique à deux variables donne donc immédiatement des informations de nature qualitatives.

Pour en tirer des informations plus quantitatives, il nous faut poser le problème de l'ajustement.

Le problème de l'établissement d'une relation fonctionnelle entre les deux séries est le problème de l'ajustement.

Dans le cas d'un nuage de points de forme **allongée**, il est possible de remplacer ce nuage par une droite **appelée droite d'ajustement affine**.

Pour tracer cette droite, on utilise **les méthodes suivants:**

### Méthode de Mayer

**Exemple 2.4.1** *Le tableau suivant donne l'évolution du nombre d'adhérents d'un club de rugby de 2001 à 2006.*

| *Année* | *2001* | *2002* | *2003* | *2004* | *2005* | *2006* |
|---|---|---|---|---|---|---|
| *Rang* $x_i$ | *1* | *2* | *3* | *4* | *5* | *7* |
| *Nombre d'adhèrents* $y_i$ | *70* | *90* | *115* | *140* | *170* | *220* |

*Le but est d'étudier cette série statistique à deux variables (le rang et le nombre d'adhérents) afin de prévoir l'évolution du*

*nombre d'adhérents pour les années suivantes.*

1. *Dans le plan muni d'un repère orthogonal d'unités graphiques : 2 cm pour une année sur l'axe des abscisses et 1 cm pour 20 adhérents sur l'axe des ordonnées, représenter le nuage de points associé à la série* $(x_i, y_i)$


2. *Calculer des coordonnées des points moyens* $G_1$ *et* $G_2$ :

   *On partage le nuage en deux groupes de même importance suivant les valeurs croissantes de* $x_i$*, et on calcule les coordonnées* $G_1; G_2$ *de chaque groupe de points:*

   $G_1$ *des années allant de 2001 à 2003,*

   $G_2$ *des années allant de 2004 à 2006,*

   $G_1\left(\frac{1+2+3}{3}; \frac{70+90+115}{3}\right), G_1\left(2; 91.7\right)$

   $G_2\left(\frac{4+5+6}{3}; \frac{140+170+220}{3}\right), G_2\left(5; 176.7\right)$

3. *La droite* $(D_1) : y = ax + b$ *la droite d'ajustement affine de Mayer qui passe par les deux point* $G_1; G_2$*.*

$a = \frac{176.7-91.7}{5-2} = 28.3$

$y_{G_1} = ax_{G_1} + b \Longleftarrow b = 91.7 - 28.3 \times 2 = 35.1$

$(D_1) : y = 28.3x + 35.1$

*Pour tracer* $(D_1)$ *, il suffit de placer* $G_1$ *et* $G_2$ *puis de tracer la droite qui les relie.*

**Méthode des moindres carrées** (**droite de régression**)

Il s'agit d'obtenir une droite équidistante des points situés de part et d'autre d'elle-même.

Pour réaliser ceci, on cherche à minimiser la somme des distances des points à la droite au carré.

On considère une série statistique à deux variables représentée par un nuage justifiant un ajustement affine.

Cette droite est **appelée droite de régression de x en y.**

**Définition 2.4.1** *La droite de régression* $D$ *de* $y$ *en* $x$ *a pour équation* $y = ax + b$ *où*

$$\begin{aligned} a &= \frac{COV(X,Y)}{V(X)} = \frac{\sigma_{xy}}{\sigma_x^2} \\ b &= \overline{y} - a\overline{x} \end{aligned}$$

**Exemple 2.4.2** *Déterminer une équation de la droite d'ajustement* $(D_2)$ *de* $y$ *en* $x$ *obtenue par la méthode des moindres carrés et la tracer*

*sur le graphique précédent.*

$\Longrightarrow (D_2)$ *:* $y = ax + b$ *avec* $a = 29$ *et* $b = 32,7$.

$(D_2)$ *:* $y = 29x + 32.7$

$\Longrightarrow$ *Tracer*

| $x$ | 0 | 8 |
|---|---|---|
| $y$ | 32.7 | 264.7 |

**Comparaison**

**Exemple 2.4.3** *En supposant que les ajustements restent valables pour les années suivantes, donner une estimation du nombre d'adhérents*

*en 2007 suivant les deux méthodes.*

$\Longrightarrow$*Dans tous les cas, il faut calculer y lorsque x correspond à l'année* 2007*, c'est à dire au rang* 7.

- *Méthode de Mayer :* $y = 28,3 \times 7 + 35,1 = 233,2$ *soit environ* 233 *adhérents .*
- *Ajustement affine :* $y = 29 \times 7 + 32,7 = 235,7$ *soit environ* 236 *adhérents .*

### 2.4.2 Coefficient de corrélation linéaire

La quantité

$$r_{xy} = \frac{COV\,(X,Y)}{\sigma_x \sigma_y}$$

s'appelle le coefficient de corrélation.

**Remarque 2.4.1**

- *Le coefficient de corrélation linéaire* $r_{xy}$ *est entre* $[-1,1]$ *, ou encore:*

$$|r_{xy}| \leq 1; -1 \leq r_{xy} \leq 1.$$

- *Plus le coefficient de régression linéaire est proche de* 1*en valeur absolue, meilleur est l'ajustement linéaire.*
- *Lorque* $r = \pm 1$*, la droite de régression passe par tous les points du nuage, qui sont donc alignés.*
- *Le coefficient de corrélation linéaire* $r_{xy}$ *permet de justifier le faire de l'ajustement linéaire.*

$$-0.7 \prec r_{xy} \prec 0.7 \Longrightarrow n'est \;\; pas \; justifié$$

$$0.7 \leq r_{xy} \leq 1 \;\; et \; -1 \leq r_{xy} \leq -0.7 \Longrightarrow est \; justifié$$

$$r_{xy} = \pm 1 \Longrightarrow parfaite$$

**Module manager: Samiha aichouche.**