X \ Y	$y_1 = 3$	$y_2 = 4$	$y_3 = 5$	*Total*
$x_1 = 2$	2	4	12	18
$x_2 = 3$	4	8	24	36
Total	6	12	36	54

Les deux variables X et Y sont indépendantes car :

1	$\frac{n_{1.} \times n_{.1}}{N} = \frac{18 \times 6}{54} = 2 = n_{11}$	4	$\frac{n_{2.} \times n_{.1}}{N} = \frac{36 \times 6}{54} = 4 = n_{21}$
2	$\frac{n_{1.} \times n_{.2}}{N} = \frac{18 \times 12}{54} = 4 = n_{12}$	5	$\frac{n_{2.} \times n_{.2}}{N} = \frac{36 \times 12}{54} = 8 = n_{22}$
3	$\frac{n_{1.} \times n_{.3}}{N} = \frac{18 \times 36}{54} = 12 = n_{13}$	6	$\frac{n_{2.} \times n_{.3}}{N} = \frac{36 \times 36}{54} = 24 = n_{23}$

3.2 Paramètres de liaison

Dans toute la suite on considère N observations sur les deux variables X et Y.

3.2.1 Covariance entre X et Y

La covariance est égale à la moyenne des écarts des couples les (x_i, y_i) de X et **Y** par rapport au point $(\bar{x}, \bar{y})$.

$$Cov(x, y) = \frac{1}{N}\sum_{i=1}^{i=N}(x_i - \bar{x})(y_i - \bar{y})$$

- **Le rôle de la covariance**

La covariance indique le sens de la relation entre les variables X et Y. Ainsi, On peut distinguer les cas suivants :

1^er^ Cas : Si $Cov(x, y) > 0$, alors on peut dire que **la relation entre les deux variables est positive**. Dans ce cas, **ces deux variables varient dans le même sens.**

2^eme^ Cas : Si $Cov(x, y) < 0$, alors on peut dire que **la relation entre les deux variables est négative**. Dans ce cas, **ces deux variables varient en sens inverse.**

3^eme^ Cas : Si $Cov(x, y) = 0$, alors on peut dire qu'**il n'y a pas de relation entre les deux variables.** Dans ce cas, **les variations de l'une n'entraînent pas la variation de l'autre.**

3.2.2 Les propriétés de la covariance

1ere Propriété

$$Cov(ax+b, cy+d) = ac.Cov(x, y)$$

Démonstration : En effet,

$$Cov(ax+b, cy+d) = \frac{1}{N}\sum_{i=1}^{i=N}\left[(ax_i+b)-(a\bar{x}+b)\right].\left[(cy_i+d)-(c\bar{y}+d)\right]$$

$$= \frac{1}{N}\sum_{i=1}^{i=N}(ax_i - a\bar{x})(cy_i - c\bar{y})$$

$$= \frac{1}{N}\sum_{i=1}^{i=N}(a\times c)(x_i-\bar{x})(y_i-\bar{y})$$

$$= (a\times c)\left[\frac{1}{N}\sum_{i=1}^{i=N}(x_i-\bar{x})(y_i-\bar{y})\right]$$

$$\Rightarrow Cov(ax+b, cy+d) = ac.Cov(x, y)$$

2eme Propriété

$$Cov(y, x) = Cov(x, y)$$

Démonstration : En effet,

$$Cov(x, y) = \frac{1}{N}\sum_{i=1}^{i=N}(x_i-\bar{x})(y_i-\bar{y}) = \frac{1}{N}\sum_{i=1}^{i=N}(y_i-\bar{y})(x_i-\bar{x})$$

$$\Rightarrow Cov(y, x) = Cov(x, y)$$

3eme Propriété

$$Cov(x, x) = V(x)$$

Démonstration : En effet,

$$Cov(x, x) = \frac{1}{N}\sum_{i=1}^{i=N}(x_i-\bar{x})(x_i-\bar{x}) = \frac{1}{N}\sum_{i=1}^{i=N}(x_i-\bar{x})^2 = V(x)$$

4eme Propriété

$$Cov(x, y) = \left(\frac{1}{N}\sum_{i=1}^{i=N}x_i y_i\right) - (\bar{x}.\bar{y})$$

Démonstration : En effet,

$$Cov(x,y)=\frac{1}{N}\sum_{i=1}^{i=N}(x_i-\bar{x})(y_i-\bar{y})=\frac{1}{N}\sum_{i=1}^{i=N}(x_iy_i-\bar{y}x_i-\bar{x}y_i+\bar{x}.\bar{y})$$

$$=\frac{1}{N}\sum_{i=1}^{i=N}(x_iy_i)-\frac{1}{N}\sum_{i=1}^{i=N}(\bar{y}x_i)-\frac{1}{N}\sum_{i=1}^{i=N}(\bar{x}y_i)+\frac{1}{N}\sum_{i=1}^{i=N}(\bar{x}.\bar{y})$$

$$=\frac{1}{N}\left(\sum_{i=1}^{i=N}(x_iy_i)-\bar{y}\underbrace{\sum_{i=1}^{i=N}(x_i)}_{N.\bar{x}}-\bar{x}\underbrace{\sum_{i=1}^{i=N}(y_i)}_{N.\bar{y}}+\underbrace{\sum_{i=1}^{i=N}(\bar{x}.\bar{y})}_{N.\bar{x}.\bar{y}}\right)$$

$$=\frac{1}{N}\left(\sum_{i=1}^{i=N}(x_iy_i)-N.\bar{x}.\bar{y}-N.\bar{x}.\bar{y}+N.\bar{x}.\bar{y}\right)$$

$$\Rightarrow Cov(x,y)=\left(\frac{1}{N}\sum_{i=1}^{i=N}x_iy_i\right)-(\bar{x}.\bar{y}).$$

3.2.3 Le coefficient de corrélation linéaire entre X et Y

Le coefficient de corrélation linéaire entre X et Y est

$$r_{x,y}=\frac{Cov(x,y)}{\sigma(x)\sigma(y)}=\frac{Cov(x,y)}{\sqrt{V(x)V(y)}}$$

Remarque 3.2.1 Le coefficient de corrélation linéaire est un nombre sans dimension car :

$$r_{x,y}=\frac{Cov(x,y)}{\sigma(x)\sigma(y)} \quad \text{et} \quad Cov(x,y)=\frac{1}{N}\sum_{i=1}^{i=N}(x_i-\bar{x})(y_i-\bar{y})$$

3.2.4 Les propriétés du coefficient de corrélation linéaire

1ère Propriété

$$r_{ax+b,cy+d}=(Signe\ de\ a)(Signe\ de\ c).r_{x,y}$$

Démonstration : En effet,

$$r_{ax+b,cy+d}=\frac{Cov(ax+b,cy+d)}{\sqrt{V(ax+b)}\sqrt{V(cy+d)}}=\frac{(a\times c).Cov(x,y)}{|a|\sqrt{V(x)}|c|\sqrt{V(y)}}$$

$$= \frac{(a\times c)}{|a|\times|c|}\frac{Cov(x,y)}{\sqrt{V(x)}\sqrt{V(y)}}$$

$$\Rightarrow r_{ax+b,cy+d} = (Signe\ de\ a)(Signe\ de\ c).r_{x,y}$$

2^{eme} Propriété

$$r_{y,x} = r_{x,y}$$

Démonstration : En effet,

$$r_{y,x} = \frac{Cov(y,x)}{\sqrt{V(y)V(x)}} = \frac{Cov(x,y)}{\sqrt{V(x)V(y)}} = r_{x,y}$$

3^{eme} Propriété

$$r_{x,x} = 1$$

Démonstration : En effet,

$$Cov(x,y) = \left(\frac{1}{N}\sum_{i=1}^{i=N} x_i y_i\right) - (\bar{x}.\bar{y}) \Rightarrow Cov(x,x) = \left(\frac{1}{N}\sum_{i=1}^{i=N} x_i^{\ 2}\right) - \bar{x}^2$$

$$\Rightarrow Cov(x,x) = V(x)$$

$$\Rightarrow r_{x,x} = \frac{Cov(x,x)}{\sqrt{V(x)V(x)}} = \frac{V(x)}{\sqrt{V(x)V(x)}} \Rightarrow r_{x,x} = 1$$

4^{eme} Propriété

Le coefficient de corrélation linéaire est compris entre -1 et 1, c'est-à-dire :

$$-1 \leq r_{x,y} \leq +1$$

Démonstration : Admis.

- **Le rôle du coefficient de corrélation linéaire**

Le coefficient de corrélation linéaire permet de mesurer le degré ou l'intensité de la liaison linéaire entre deux variables statistiques. C'est-à-dire :

1. Si $r_{x,y} = 1$, on dit qu'il y a une parfaite corrélation linéaire positive entre les deux

variables.

2. Si $r_{x,y} = -1$, on dit qu'il y a une parfaite corrélation linéaire négative entre les deux variables.

3. Si $r_{x,y} = 0$, on dit qu'il y a absence de corrélation linéaire entre les deux variables.

4. On dit qu'il y a une forte corrélation linéaire entre les deux variables (ou forte dépendance linéaire) si $r_{x,y}$ est proche de ± 1.

5. En revanche, si $r_{x,y}$ est proche de zéro (0), on dit qu'il y a une faible corrélation linéaire entre les deux variables.

3.3 Ajustement d'un nuage de points

Dans toute la suite on considère N observations sur les deux variables X et Y.

3.3.1 Nuage de points

Ensemble de points isolés représentés dans un graphique cartésien ; c'est-à-dire des points $M_1, M_2, \dots, M_n$ de coordonnées (x_1, y_1) ; (x_2, y_2); … ; (x_n, y_n)

Exemple 3.3.1 On considère le tableau suivant, relatif à une population associé à deux variables mesurées sur 13 bébés tels que, X = « le poids du bébé » et Y = « la taille du bébé »

Masse (*kg*)	3,3	3,8	4,6	5,4	6,0	6,6	7,1	7,6	8,1	8,4	8,7	9,0	9,3
Taille (*cm*)	49,4	52,4	55,6	58,7	61,0	63,0	64,8	66,4	67,8	69,0	70,3	72,6	72,9

Le nuage des points de coordonnées (*3,3 ; 49,4*); (*3,8 ; 52,4*); (*4,6 ; 55,6*); (*5,4 ; 58,7*); (*6,0 ; 61,0*); (*6,6 ; 63,0*); (*7,1 ; 64,8*); (*7,6 ; 66,4*); (*8,1 ; 67,8*); (*8,4 ; 69,0*); (*8,7 ; 70,3*); (*9,0 ; 72,6*); (*9,3 ; 72,9*) *est :*