I. Intervalle de Confiance :

Dans une série unimodale et symétrique, on :

$I_1 = [\overline{x} - \sigma(x)\ ;\ \overline{x} + \sigma(x)]$ *avec* 68% *des valeurs*

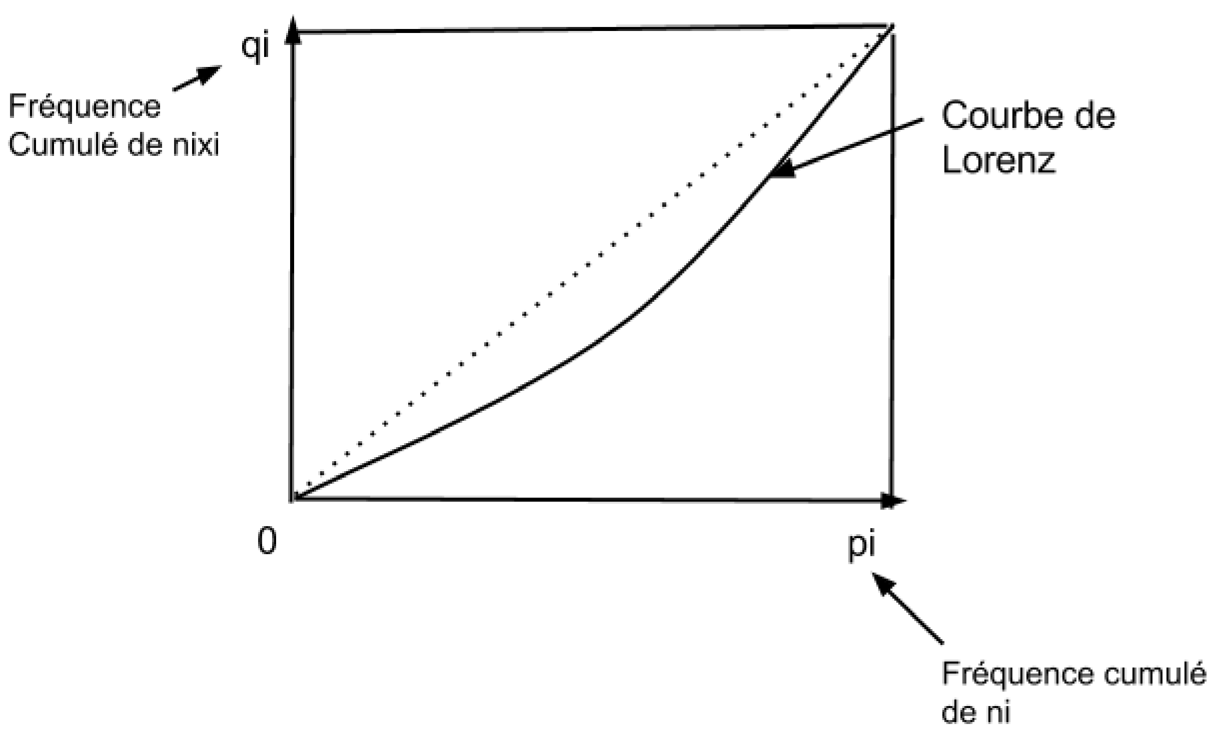
$I_2 = [\overline{x} - 2\sigma(x)\ ;\ \overline{x} + 2\sigma(x)]$ *avec* 95% *des valeurs*

$I_3 = [\overline{x} - 3\sigma(x)\ ;\ \overline{x} + 3\sigma(x)]$ *avec* 99% *des valeurs*

II. Paramètres de Concentration :

Ils montrent l'égalité dans une distribution, dans la plupart des cas, on les utilise dans le calcul de la masse salariale.

A. Courbe de Lorenz :



On calcule les "qi" en utilisant la formule suivante :

$$q_i = \frac{\sum_{j=1}^{n} n_j x_j}{\sum_{i=1}^{n} n_i x_i}$$

avec :

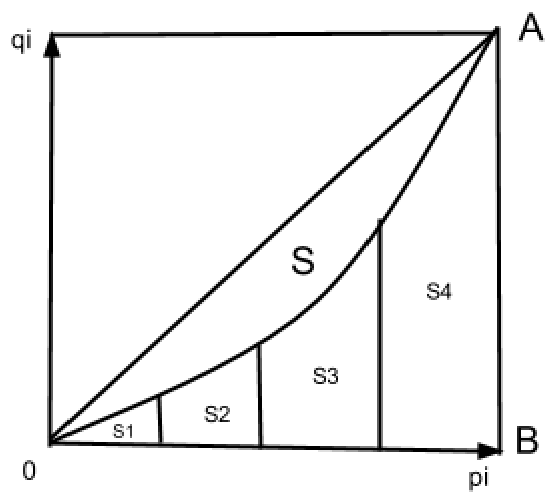
$$\sum_{i=1}^{n} n_i x_i \ : \ masse\ total$$

$$\sum_{j=1}^{n} n_j x_j \ : \ masse\ cumulé\ (nixi\ cumulé)$$

B. Indice de Gini :

On a :

$$I_G = \frac{S}{S_0}\ avec\ S_0\ l'air\ du\ triangle\ OAB$$



$$S_0 = \frac{Base \times Hauteur}{2} = \frac{1 \times 1}{2} = 0,5\ en\ fréquence\ réel\ et\ 5000\ en\ pourcentage$$

$$S = S_0 - \{S_1 + S_2 + S_3 + S_4\}$$

S1 est un triangle tandis que S2, S3, et S4 sont des trapèze, alors :

$$S = S_0 - \left\{\frac{Base \times Hauteur}{2} + \frac{Base \times Somme\ des\ Côtés}{2} + \frac{Base \times Somme\ des\ Côtés}{2} + \frac{Base \times Somme\ des\ Côtés}{2}\right\}$$

$$S = S_0 - \left\{\frac{p_1 \times q_1}{2} + \frac{(p_2 - p_1) \times (q_1 + q_2)}{2} + \frac{(p_3 - p_2) \times (q_2 + q_3)}{2} + \frac{(p_4 - p_3) \times (q_3 + q_4)}{2}\right\}$$

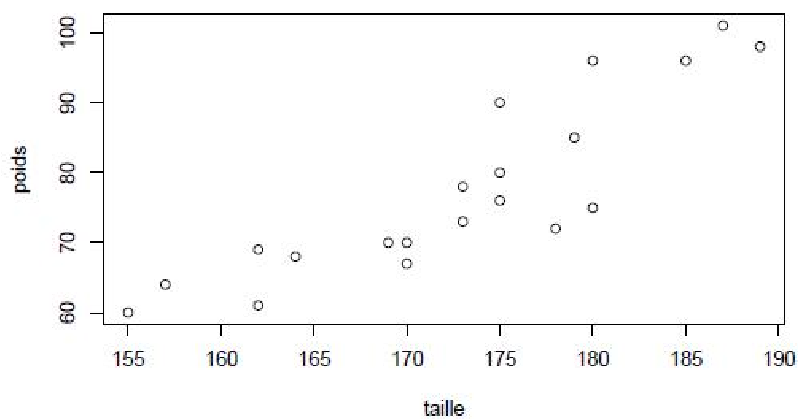
Interprétation :

- Répartition Homogène ou pas.
- Concentration forte ou faible.

III. Série à deux variables :

Représentation par le **nuage de point.**

y_i	x_i	y_i	x_i
60	155	75	180
61	162	76	175
64	157	78	173
67	170	80	175
68	164	85	179
69	162	90	175
70	169	96	180
70	170	96	185
72	178	98	189
73	173	101	187



Calcul de la **Covariance** et le **coefficient de corrélation** pour la détermination de la nature de liaison entre les deux variables.

$$Cov(x,y) = \frac{1}{n}\sum_{i=1}^{n} n_{ij}(x_i - \overline{x})(y_i - \overline{y}) = \frac{1}{n}\sum_{i=1}^{n} n_{ij} x_i y_i - \overline{x}\,\overline{y}$$

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

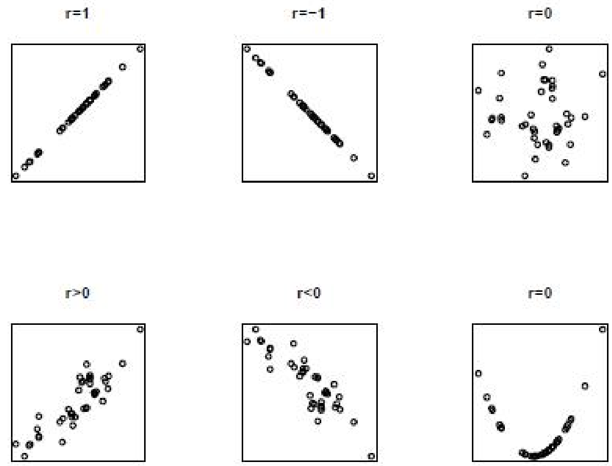
$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} n_i x_i$ $\qquad$ $V(x) = \frac{1}{n}\sum_{i=1}^{n} n_i (x_i - \overline{x})^2$

$\overline{y} = \frac{1}{n}\sum_{j=1}^{n} n_j y_j$ $\qquad$ $V(y) = \frac{1}{n}\sum_{j=1}^{n} n_j (y_j - \overline{y})^2$

Ces paramètres sont appelés **paramètres marginaux** : variances marginales, moyennes marginales, écarts-types marginaux, quantiles marginaux, etc.

Remarque : Le coefficient de corrélation mesure la **dépendance linéaire** entre deux variables :

$$-1 \leq r_{xy} \leq 1$$



Méthode des **moindres carrées** :

$$D_{y/x} \;:\; y = ax + b$$

$$a = \frac{Cov(x,y)}{V(x)} \qquad\qquad b = \overline{y} - a\overline{x}$$

$$D_{x/y} \;:\; x = a'y + b'$$

$$a' = \frac{Cov(x,y)}{V(y)} \quad b = \overline{x} - a\overline{y}$$