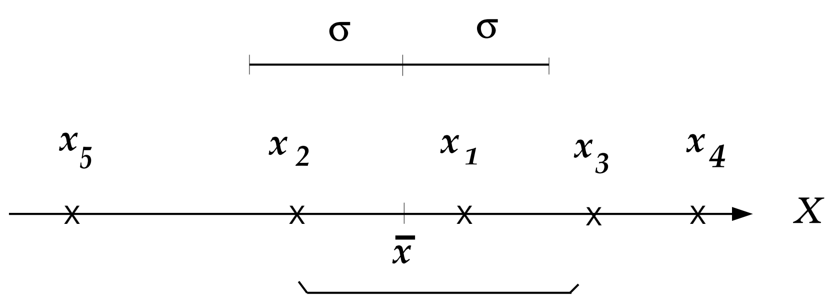
5 REPRESENTATION GEOMETRIQUE DE LA DISPERSION
Une variable



$(x-\overline{x})^2 \leq \sigma^2 \Rightarrow$ segment de longueur $2.\sigma$

$$\Leftrightarrow \left(\frac{x-\overline{x}}{\sigma}\right)^2 \leq 1$$

5 REPRESENTATION GEOMETRIQUE DE LA DISPERSION

Une variable

- On a vu que pour une variable, σ est un paramètre de dispersion.

- On cherche une équation permettant de caractériser l'appartenance d'un point x_i au segment

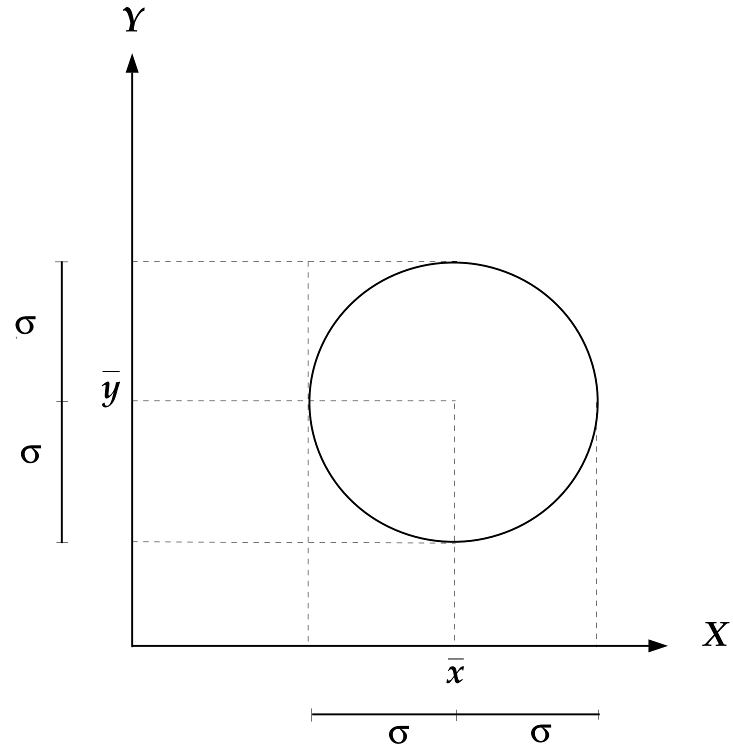
$$[\overline{x} - \sigma, \overline{x} + \sigma]$$

- On établit : $\boxed{(x - \overline{x})^2 \leq \sigma^2}$

5 REPRESENTATION GEOMETRIQUE DE LA DISPERSION
Deux variables

X et Y avec $\sigma_x = \sigma_y = \sigma$ et $\rho = 0$

$(x - \overline{x})^2 + (y - \overline{y})^2 \leq \quad \sigma^2 \Rightarrow$ disque de rayon σ



5 REPRESENTATION GEOMETRIQUE DE LA DISPERSION
Deux variables

$$(x - \overline{x})^2 \leq \sigma^2$$

permet de passer à 2 variables avec :

$$\sigma_x = \sigma_y = \sigma \;\text{ et }\; \rho = 0$$

$$\boxed{(x - \overline{x})^2 + (y - \overline{y})^2 \leq \sigma^2}$$

qui est un disque de rayon σ.

5 REPRESENTATION GEOMETRIQUE DE LA DISPERSION

Ecriture matricielle

$$\left(\frac{x-\overline{x}}{\sigma}\right)^2 + \left(\frac{y-\overline{y}}{\sigma}\right)^2 \leq 1$$

$$\Leftrightarrow \left[\left(\frac{x-\overline{x}}{\sigma}\right) \ \left(\frac{y-\overline{y}}{\sigma}\right)\right] . \begin{bmatrix} \left(\frac{x-\overline{x}}{\sigma}\right) \\ \left(\frac{y-\overline{y}}{\sigma}\right) \end{bmatrix} \leq 1$$

$$\Leftrightarrow \left[(x-\overline{x}) \ (\mathrm{y}-\overline{\mathrm{y}})\right] . \begin{bmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{1}{\sigma^2} \end{bmatrix} . \begin{bmatrix} (x-\overline{x}) \\ (\mathrm{y}-\overline{\mathrm{y}}) \end{bmatrix} \leq 1$$

$$\Leftrightarrow \left[(x-\overline{x}) \ (\mathrm{y}-\overline{\mathrm{y}})\right] . \begin{bmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{bmatrix}^{-1} . \begin{bmatrix} (x-\overline{x}) \\ (\mathrm{y}-\overline{\mathrm{y}}) \end{bmatrix} \leq 1$$

$$\Leftrightarrow V^{'} . \Sigma^{-1} . V \leq 1$$

5 REPRESENTATION GEOMETRIQUE DE LA DISPERSION

Ecriture matricielle

Cette équation du disque :

$$\boxed{(x-\overline{x})^2+(y-\overline{y})^2 \leq \sigma^2}$$

peut aussi se mettre sous la forme :

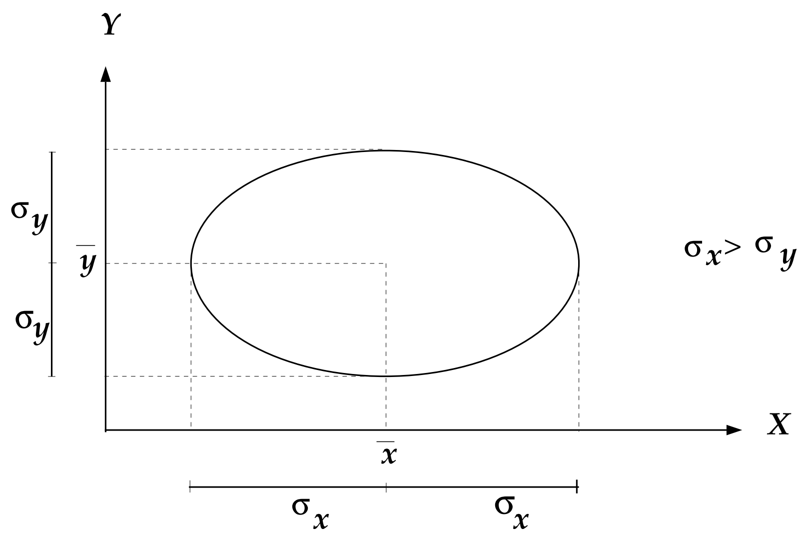
$$\boxed{V^{'}.\Sigma^{-1}.V \leq 1}$$

$$\text{avec } \ \Sigma = \begin{bmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{bmatrix}$$

$$\text{et } \ V = \begin{bmatrix} (x-\overline{x}) \\ (y-\overline{y}) \end{bmatrix}$$

5 REPRESENTATION GEOMETRIQUE DE LA DISPERSION

Deux variables : X et Y avec $\sigma_X \neq \sigma_Y$ et $\rho = 0$



Equation de l'ellipsoïde de dispersion associée à Σ.

$$V.\Sigma^{-1}.V \leq 1$$

$$\Leftrightarrow \left[(x-\overline{x}), (y-\overline{y})\right] . \begin{bmatrix} \sigma_x^2 & 0 \\ 0 & \sigma_y^2 \end{bmatrix}^{-1} . \begin{bmatrix} (x-\overline{x}) \\ (y-\overline{y}) \end{bmatrix} \leq 1$$

5 REPRESENTATION GEOMETRIQUE DE LA DISPERSION

Deux variables : X et Y avec $\sigma_X \neq \sigma_Y$ et $\rho = 0$

Soit : $\Sigma = \begin{bmatrix} \sigma_x^2 & 0 \\ 0 & \sigma_y^2 \end{bmatrix}$ la matrice de variance-covariance.

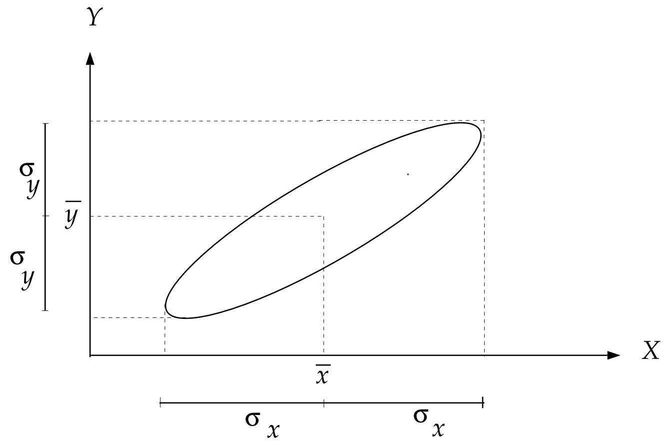
L'équation de l'ellipse est :

$$\boxed{V^{'}.\Sigma^{-1}.V \leq 1}$$

Remarque : Le grand axe de l'ellipse est dans le sens du maximum entre σ_x et σ_y.

5 REPRESENTATION GEOMETRIQUE DE LA DISPERSION

Deux variables : X et Y avec $\sigma_X \neq \sigma_Y$ et $\rho \neq 0$



Equation de l'ellipsoïde de dispersion associée à la forme quadratique $V'.\Sigma.V$

$$V^{'}.\Sigma^{-1}.V \leq 1$$

$$\Leftrightarrow [(x-\overline{x}),(y-\overline{y})] \cdot \begin{bmatrix} \sigma_x^2 & \rho.\sigma_x.\sigma_y \\ \rho.\sigma_x.\sigma_y & \sigma_y^2 \end{bmatrix}^{-1} \cdot \begin{bmatrix} (x-\overline{x}) \\ (y-\overline{y}) \end{bmatrix} \leq 1$$

5 REPRESENTATION GEOMETRIQUE DE LA DISPERSION

Deux variables : X et Y avec $\sigma_X \neq \sigma_Y$ et $\rho \neq 0$
Soit :

$$\Sigma = \begin{bmatrix} \sigma_x^2 & \rho.\sigma_x.\sigma_y \\ \rho.\sigma_x.\sigma_y & \sigma_y^2 \end{bmatrix}$$

la matrice de variance-covariance.

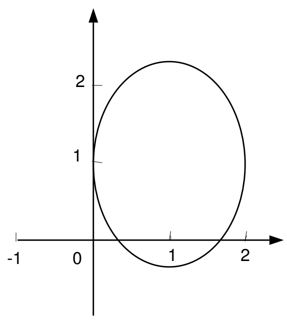
L'équation de l'ellipse est :

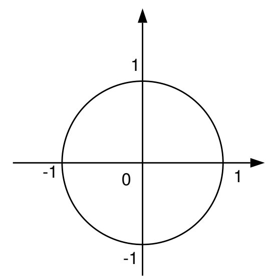
$$\boxed{V^{'}.\Sigma^{-1}.V \leq 1}$$

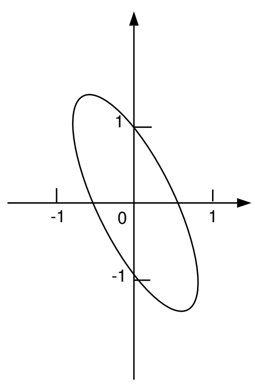
Remarque : Ellipse oblique car $\rho \neq 0$.

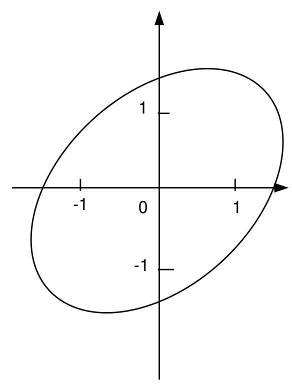
Exemples

$$\mu=(1,1), \Sigma=\begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \qquad \mu=(0,0)\,, \Sigma=\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$









$$\mu=(0,0)\,, \Sigma=\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \qquad \mu=(0,0)\,, \Sigma=\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

Exemples

Ces ellipses nous donnent une première idée de la forme du nuage de points et de la structure de la matrice Σ.

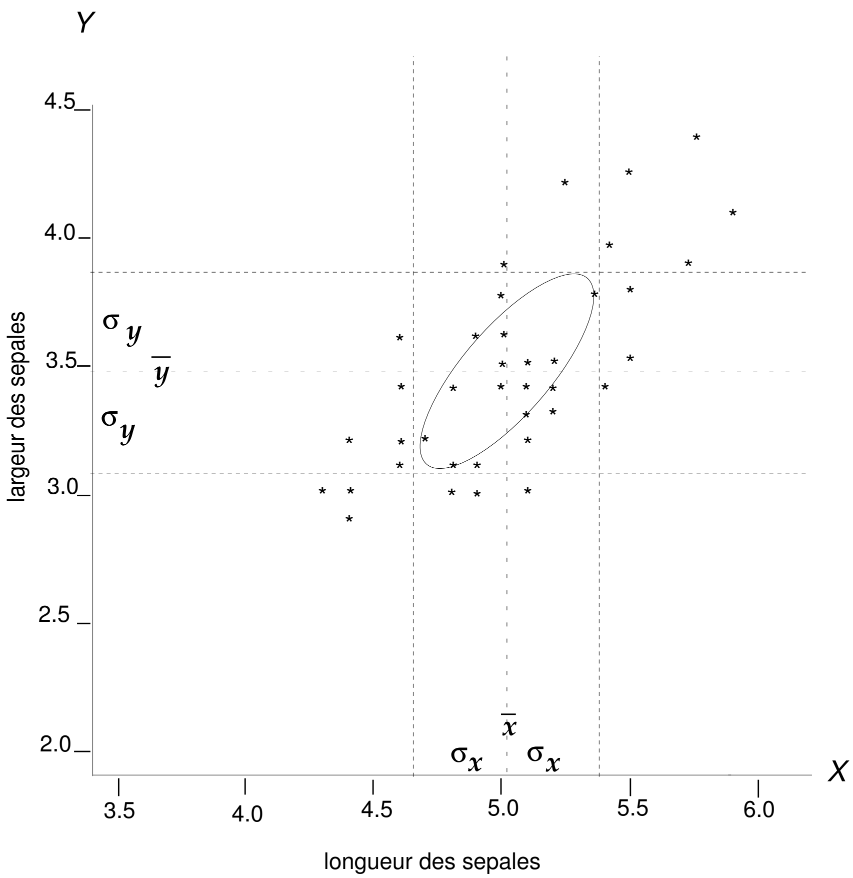
Remarques :

- Le cercle correspond à une matrice Σ très particulière : une matrice diagonale avec une même valeur sur la diagonale.

$$\forall_i \qquad \sigma_i = \sigma \quad \text{et} \quad \rho = 0$$

- Tout ceci est généralisable à plus de 2 variables.

Exemple : iris *setosa*



Equation de l'ellipse : $V^{'}.\Sigma.V \leq 1$

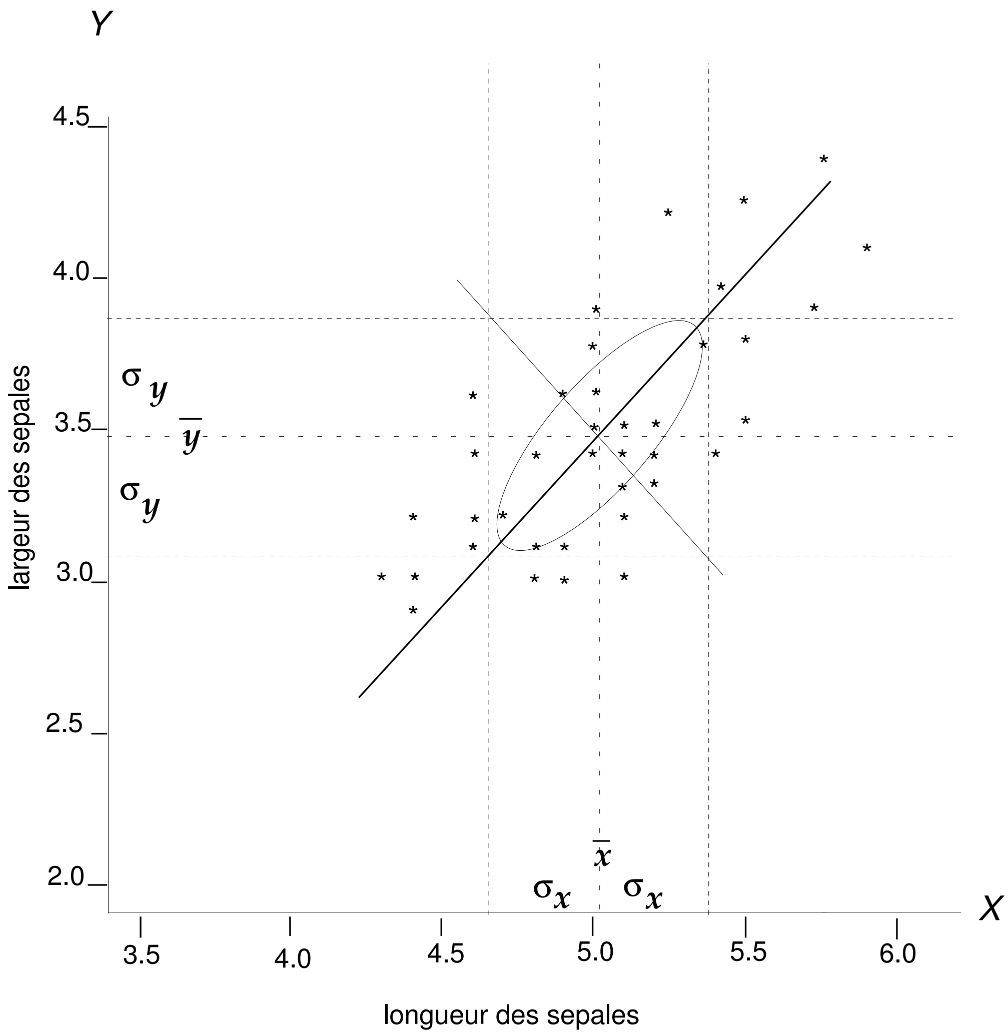
Exemple : Iris *setosa*

On considère l'ellipse de dispersion associée à la forme quadratique : $V^{'}.\Sigma.V$ où $\Sigma = \begin{bmatrix} 0,124 & 0,099 \\ 0,099 & 0,144 \end{bmatrix}$.

- Bien faire remarquer que pour l'axe des X, l'ellipse est dans l'intervalle : $[\overline{x} - \sigma_x, \overline{x} + \sigma_x]$ c'est-à-dire que la projection orthogonale de l'ellipse sur l'axe X donne ce segment. Il en est de même pour l'axe des Y.

- On peut aussi remarquer que le grand axe le l'ellipse est dans la direction de plus grande variabilité du nuage de points. A l'inverse, le petit axe de l'ellipse correspond à la direction de plus faible variabilité.

Exemple : *Iris setosa*

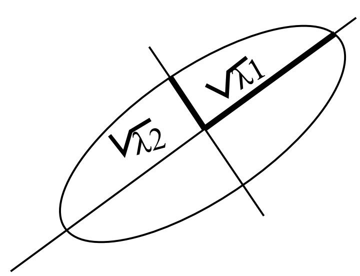


Exemple : Iris *setosa*

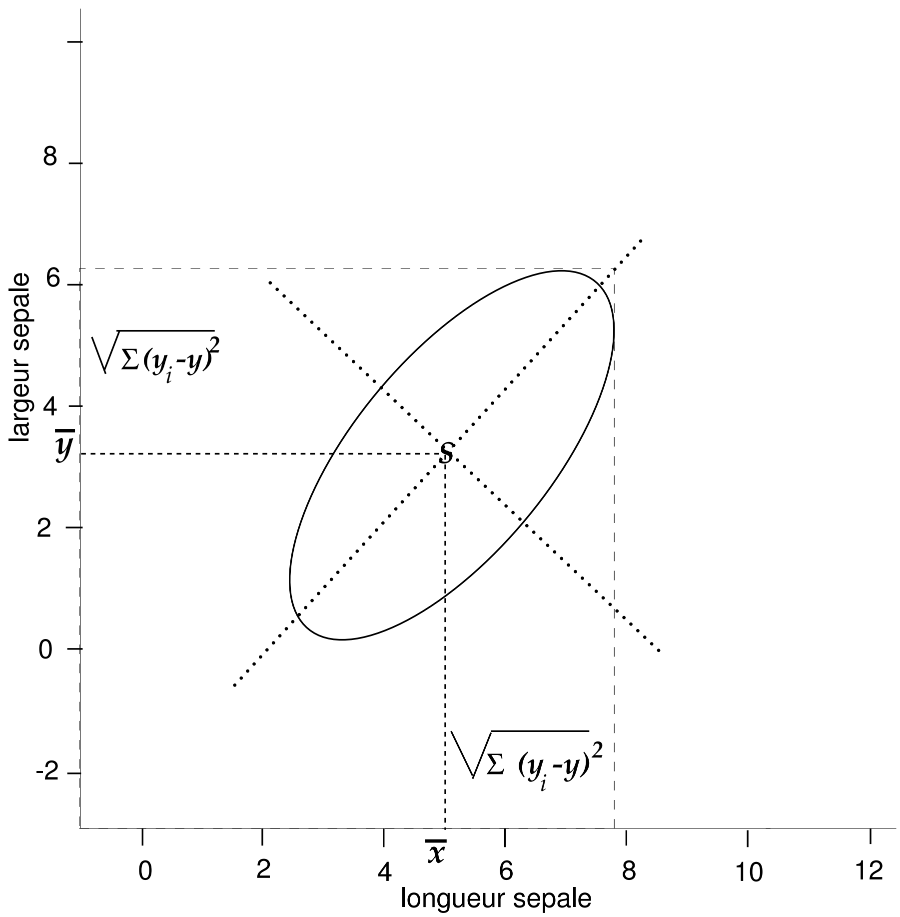
On a tracé ici le grand axe et le petit axe.

- Le plus grand axe de l'ellipse (+ grande variabilité) est la direction du vecteur propre associé à la plus grande valeur propre λ_1 de Σ.
- De même le petit axe est la direction du vecteur propre associé à la plus petite valeur propre λ_2 de Σ.

  Les vecteurs et valeurs propres d'une matrice sont donc des outils intéressants.
- De plus la longueur du 1/2 grand axe de l'ellipse $= \sqrt{\lambda 1}$ et la longueur du 1/2 petit axe $= \sqrt{\lambda 2}$ avec $\lambda_1 > \lambda_2$.



Exemple : *Iris setosa*



Exemple : iris *setosa*

L'ellipse d'inertie associée à W ou à la forme quadratique $V'.W.V$ a pour équation : $V'.W^{-1}.V \leq 1$.

Par ailleurs, la matrice de variance-covariance Σ est liée à la matrice d'inertie W par la relation :

$$\Sigma = \frac{1}{n}.W$$

- On obtient donc des résultats analogues pour Σ ou pour W.
- Habituellement, on utilise la matrice W des sommes de carrés plutôt que la matrice Σ des valeurs de W moyennées par n.