# 27 Statistiques bivariées

ECG Maths Appliquées
Semestre 4

On aborde dans cette partie le cas où l'on observe non plus un, mais deux caractères $x$ et $y$ des individus de la population $\Omega$, celle-ci étant formée d'individus $\omega_1, \dots, \omega_n$.

**Définition 27.1**

On appelle **série statistique bivariée** (ou double) sur $\Omega$ toute application $(x, y)$ de $\Omega$ dans $\mathbb{R}^2$. Une observation est donc un couple $(x_i, y_i)$ de réels ($x_1, \dots, x_n$ ne sont pas forcément distincts deux à deux, ainsi que $y_1, \dots, y_n$).
Si $\{(x_i, y_j),\ (i, j) \in [\![1, p]\!] \times [\![1, q]\!]\}$ est l'ensemble des valeurs différentes observées par $(x, y)$ et si, pour tout $(i, j) \in [\![1, p]\!] \times [\![1, q]\!]$, $n_{i,j}$ est le nombre de fois que le couple $(x_i, y_i)$ a été observé, la série sera aussi notée $((x_i, y_j), n_{i,j})_{\substack{1 \leqslant i \leqslant p \\ 1 \leqslant j \leqslant q}}$.

## A. Nuage de point, point moyen

**Définition 27.2**

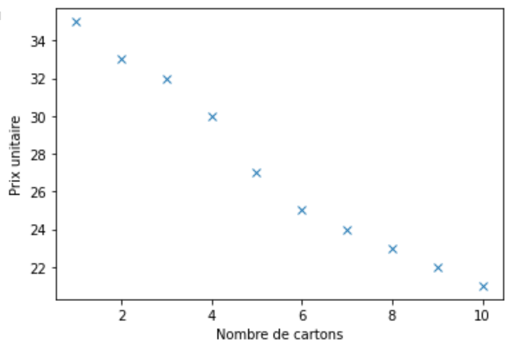
On munit le plan d'un repère orthogonal.
Étant donnée une série statistique double $((x_i, y_i))_{1 \leqslant i \leqslant n}$, l'ensemble des points $M_i$ du plan de coordonnées $(x_i, y_i)$ (pour $i \in [\![1, n]\!]$) est appelé **nuage de points** de la série.

**Exemples 27.1**

**a.** Un grossiste en papier applique un tarif dégressif en fonction de la quantité vendue. Le tableau suivant présente un extrait de ses tarifs :

| Nombre de cartons $x_i$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Prix unitaire $y_i$ en € | 35 | 33 | 32 | 30 | 27 | 25 | 24 | 23 | 22 | 21 |

Voici une représentation du nuage de point :

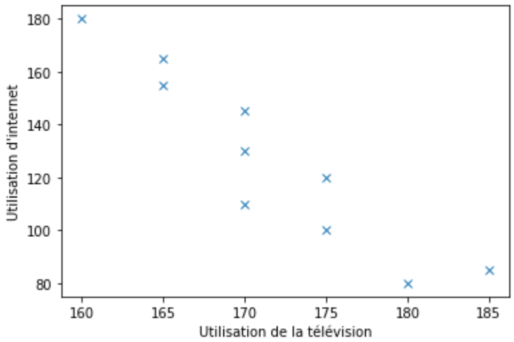



**b.** Le tableau ci-après donne des estimations du temps moyen passé par jour dans le monde devant la télévision ($x_i$) et sur internet ($y_i$), en minutes :

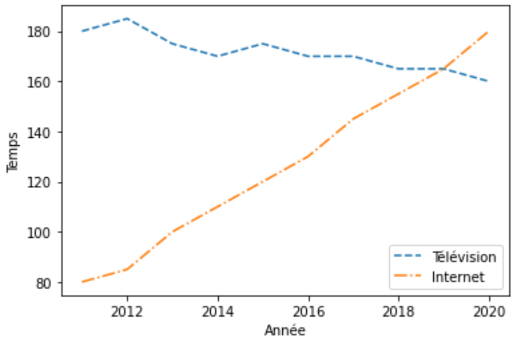
| Année | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 |
|---|---|---|---|---|---|---|---|---|---|---|
| $x_i$ | 180 | 185 | 175 | 170 | 175 | 170 | 170 | 165 | 165 | 160 |
| $y_i$ | 80 | 85 | 100 | 110 | 120 | 130 | 145 | 155 | 165 | 180 |

© www.stephanepreteseille.com

Il peut être intéressant d'étudier l'évolution des couples $(x_i, y_i)$, c'est à cela que sert la représentation du nuage de point :




Remarquons cependant qu'il est parfois utile d'analyser cette série double en étudiant chacune des séries séparément :




**Définition 27.3**

On appelle **point moyen** du nuage de points d'une série statistique double $((x_i, y_i))_{1 \leqslant i \leqslant n}$ le point du plan de coordonnées $(\overline{x}, \overline{y})$.

**Exemple 27.2** Dans le premier exemple de 23.11, le point moyen de la série est le point de coordonnées $(5.5, 27.2)$.

## B. Description d'une série statistique bivariée

**Définition 27.4**

Soit $((x_i, y_j, n_{i,j}))_{\substack{1 \leqslant i \leqslant p \\ 1 \leqslant j \leqslant q}}$ une série statistique double.

*i.* On appelle **effectif total** de la série le nombre $n = \sum_{i=1}^{p} \sum_{j=1}^{q} n_{i,j}$.

*ii.* On appelle **fréquence** de $(x_i, y_j)$ le nombre $f_{i,j} = \frac{n_{i,j}}{n}$.

*iii.* On appelle **effectif marginal de $x$ en $x_i$** le nombre $\boldsymbol{n_{i,\cdot} = \sum_{j=1}^{q} n_{i,j}}$.

*iv.* On appelle **effectif marginal de $y$ en $y_j$** le nombre $\boldsymbol{n_{\cdot,j} = \sum_{i=1}^{p} n_{i,j}}$.

*v.* La série statistique simple $\boldsymbol{((x_i, n_{i,\cdot}))_{1 \leqslant i \leqslant p}}$ est appelée **série marginale en $x$**.

*vi.* La série statistique simple $\boldsymbol{((y_i, n_{\cdot,j}))_{1 \leqslant j \leqslant q}}$ est appelée **série marginale en $y$**.

© www.stephanepreteseille.com. Diffusion et reproduction interdites, même partiellement

**Remarques**

**a.** Comme dans le cas de séries statistiques simples, il arrive que les éléments d'une série statistique double soient groupés en classe.On parle alors de série statistique double continue $(([a_i, a_{i+1}[, [b_j, b_{j+1}[, n_{i,j}))_{\substack{1 \leqslant i \leqslant p \\ 1 \leqslant j \leqslant q}}$. On adopte dans ce cas des définitions analogues à celles des séries doubles discrètes, les calculs se faisant alors avec le centre de chaque classe.

**b.** Les informations sur les effectifs et les fréquences d'une série statistique double sont en général représentées sous forme de tableaux à double entrées :

| **Effectifs** | $y_1$ | $\cdots$ | $y_j$ | $\cdots$ | $y_q$ | Somme |
|---|---|---|---|---|---|---|
| $x_1$ | $n_{1,1}$ | $\cdots$ | $n_{1,j}$ | $\cdots$ | $n_{1,q}$ | $n_{1,\cdot}$ |
| $\vdots$ | $\vdots$ | $\ldots$ | $\vdots$ | $\ldots$ | $\vdots$ | $\vdots$ |
| $x_i$ | $n_{i,1}$ | $\cdots$ | $n_{i,j}$ | $\cdots$ | $n_{i,q}$ | $n_{i,\cdot}$ |
| $\vdots$ | $\vdots$ | $\ldots$ | $\vdots$ | $\ldots$ | $\vdots$ | $\vdots$ |
| $x_p$ | $n_{p,1}$ | $\cdots$ | $n_{p,j}$ | $\cdots$ | $n_{p,q}$ | $n_{p,\cdot}$ |
| Somme | $n_{\cdot,1}$ | $\cdots$ | $n_{\cdot,j}$ | $\cdots$ | $n_{\cdot,q}$ | $n$ |

| **Fréquences** | $y_1$ | $\cdots$ | $y_j$ | $\cdots$ | $y_q$ | Somme |
|---|---|---|---|---|---|---|
| $x_1$ | $f_{1,1}$ | $\cdots$ | $f_{1,j}$ | $\cdots$ | $f_{1,q}$ | $f_{1,\cdot}$ |
| $\vdots$ | $\vdots$ | $\ldots$ | $\vdots$ | $\ldots$ | $\vdots$ | $\vdots$ |
| $x_i$ | $f_{i,1}$ | $\cdots$ | $f_{i,j}$ | $\cdots$ | $n_{i,q}$ | $f_{i,\cdot}$ |
| $\vdots$ | $\vdots$ | $\ldots$ | $\vdots$ | $\ldots$ | $\vdots$ | $\vdots$ |
| $x_p$ | $f_{p,1}$ | $\cdots$ | $f_{p,j}$ | $\cdots$ | $f_{p,q}$ | $f_{p,\cdot}$ |
| Somme | $f_{\cdot,1}$ | $\cdots$ | $f_{\cdot,j}$ | $\cdots$ | $f_{\cdot,q}$ | $n$ |

## C. Caractéristiques d'une série statistique bivariée

**Définition 27.5**

Soit $((x_i, y_j), n_{i,j})_{\substack{1 \leqslant i \leqslant p \\ 1 \leqslant j \leqslant q}}$ une série statistique double. On appelle **covariance** de la série le réel

$$\mathrm{Cov}(x, y) = \frac{1}{n} \sum_{i=1}^{p} \sum_{j=1}^{q} n_{i,j} \left(x_i - \overline{x}\right) \left(y_j - \overline{y}\right)$$

**Remarques**

**a.** Si $x$ ou $y$ ne prend qu'une seule valeur, alors $\mathrm{Cov}(x, y) = 0$.

**b.** Les calculs sont souvent fastidieux, mais ils peuvent être effectués à l'aide de Python par exemple.

**c.** Le plus souvent, on dispose d'une série statistique double $((x_i, y_i))_{1 \leqslant i \leqslant n}$ constituées de relevés et dont les données sont donc listées et se répètent autant de fois que leur effectif. Dans ce cas on a $p = q = n$ et :

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

www.stephanepreteseille.com ©

© www.stephanepreteseille.com. Diffusion et reproduction interdites, même partiellement

Le calcul de la covariance à l'aide de Python peut donc se faire à l'aide de la fonction suivante :

```
import numpy as np
def cov(x,y):
    n=len(x)
    moy_x=np.mean(x)
    moy_y=np.mean(y)
    s=0
    for i in range(0,n):
        s+=(x[i]-moy_x)*(y[i]-moy_y)
    return s/n
```

On pourra se souvenir que la bibliothèque `numpy` offre une fonction `cov` permettant de calculer la covariance de $(x, y)$ : l'instruction `np.cov(x,y,bias=True)` renvoie ainsi la matrice suivante (appelée matrice de variance-covariance) :

$$\begin{pmatrix} \mathbb{V}(x) & \mathrm{Cov}(x,y) \\ \mathrm{Cov}(x,y) & \mathbb{V}(y) \end{pmatrix}$$

Ainsi l'instruction `np.cov(x,y,bias=True)[0,1]` permet de renvoyer la covariance de $(x, y)$.

**d.** Plus généralement, on peut calculer la covariance d'une série statistique double groupée $((x_i, y_j, n_{i,j}))_{\substack{1\leqslant i\leqslant p\\ 1\leqslant j\leqslant q}}$ à l'aide de la fonction suivante (où l'on suppose que `x` et `y` sont des vecteurs lignes contenant les données $x$ et $y$ et que `n` est une matrice dont le coefficient `n[i,j]` correspond à l'effectif de $(x_i, y_j)$ :

```
import numpy as np
def cov(x,y,n):
    p=len(x)
    q=len(y)
    moy_x=np.mean(x)
    moy_y=np.mean(y)
    s=0
    for i in range(0,p):
        for j in range(0,q):
            s+=n[i,j]*(x[i]-moy_x)*(y[j]-moy_y)
    return s/np.sum(n)
```

**Exemple 27.3** Reprenons la première série proposée dans l'exemple 23.11. Pour calculer la covariance de $(x, y)$, on peut par exemple utiliser la fonction définie dans le point c de la remarque précédente ; les commandes suivantes dans le programme principal :

```
x=np.arange(0,10)
y=np.array([35,33,32,30,27,25,24,23,22,21])
n=np.eye(10)
print("cov(x,y)=",cov(x,y))
```

renvoient le résultat suivant :

$$\mathrm{Cov}(x, y) = -13.4$$

© www.stephanepreteseille.com. Diffusion et reproduction interdites, même partiellement

**Théorème 27.6 ► Formule de Koenig-Huygens**

Soit $((x_i, y_j, n_{i,j}))_{\substack{1\leqslant i\leqslant p\\ 1\leqslant j\leqslant q}}$ une série statistique double. On a :

$$\operatorname{Cov}(x,y) = \frac{1}{n}\sum_{i=1}^{p}\sum_{j=1}^{q} n_{i,j}x_iy_j - \left[\frac{1}{n}\sum_{i=1}^{p} n_{i,\cdot}x_i\right]\left[\frac{1}{n}\sum_{j=1}^{q} n_{\cdot,j}y_j\right] = \frac{1}{n}\sum_{i=1}^{p}\sum_{j=1}^{q} n_{i,j}x_iy_j - \overline{x}\cdot\overline{y}$$

**Remarque** Dans le cas où les données ne sont pas groupées, c'est-à-dire dans le cas d'une série statistique double $((x_i, y_i))_{1\leqslant i\leqslant n}$ où les couples $(x_i, y_i)$ ne sont pas deux à deux distincts, la formule devient :

$$\operatorname{Cov}(x,y) = \frac{1}{n}\sum_{i=1}^{n} x_iy_j - \overline{x}\cdot\overline{y}$$

**Définition 27.7**

Soit $((x_i, y_j, n_{i,j}))_{\substack{1\leqslant i\leqslant p\\ 1\leqslant j\leqslant q}}$ une série statistique double telle que les écarts-types des séries marginales en $x$ et en $y$ soient non nuls. On appelle **coefficient de corrélation linéaire** de la série le réel

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

**Théorème 27.8**

Soit $((x_i, y_j, n_{i,j}))_{\substack{1\leqslant i\leqslant p\\ 1\leqslant j\leqslant q}}$ une série statistique double telle que les écarts-types des séries marginales en $x$ et en $y$ soient non nuls. On a :

$$|\sigma(x,y)| \leqslant 1$$

et $|\sigma(x)| = 1$ si et seulement si les points du nuage de la série sont alignés.

**Exercice 27.1** On se propose de démontrer le théorème 27.8. Pour cela, on suppose que les variances respectives de $x$ et de $y$ sont non nulles et on considère la fonction polynôme $P$ définie par :

$$\forall t \in \mathbb{R},\ P(t) = \frac{1}{n}\sum_{i=1}^{p}\sum_{j=1}^{q} n_{i,j}\left(t\left(x_i - \overline{x}\right) - \left(y_j - \overline{y}\right)\right)^2$$

Vérifier que $P$ est une fonction polynôme de degré 2 et l'écrire sous la forme $P(t) = at^2 + bt + c$ puis en déduire une preuve du théorème 27.8.

## D. Droite de régression

Observons les deux nuages de points obtenus dans les exemples précédents. Dans le premier cas, il semble que les points soient proches d'une droite ; cette remarque peut être intéressante en termes de conjecture : que se passera-t-il si l'évolution reste proche d'une droite ? Dans ce cas, il peut donc être pertinent de déterminer l'équation d'une droite la plus proche possible des points du nuage. On parle alors d'ajustement affine. Il existe plusieurs ajustements affines possible, mais un bon ajustement affine se fera toujours par une droite passant pas le point moyen du nuage.

Une méthode pour obtenir un ajustement affine est la méthode des moindres carrés : déterminer les réels $a$ et $b$ rendant minimale la somme $\sum\limits_{i=1}^{n}\left[y_i - (ax_i + b)\right]^2$.

**Exercice 27.2** Soit $((x_i, y_i))_{1\leqslant i\leqslant n}$ une série statistique double. On suppose que $x_1, \ldots, x_n$ ne sont pas tous égaux. La répartition de ces points dans le plan indique que les points semblent alignés.

On cherche alors une droite s'ajustant le plus possible à ce nuage de points et l'on se propose pour cela de chercher un couple $(a, b)$ de réels rendant minimale la fonction

$$f : (a,b) \mapsto \sum_{i=1}^{n}\left[y_i - (ax_i + b)\right]^2$$

© www.stephanepreteseille.com. Diffusion et reproduction interdites, même partiellement

On note alors respectivement $\overline{x}$ et $\sigma_x^2$ (avec $\sigma_x \geqslant 0$) la moyenne et la variance de la série statistique $(x_i)_{1\leqslant i\leqslant n}$, c'est-à-dire :

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad \text{et} \quad \sigma_x^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \overline{x}^2$$

On note de même respectivement $\overline{y}$ et $\sigma_y^2$ la moyenne et la variance de la série statistique $(y_i)_{1\leqslant i\leqslant n}$.
Enfin, on note $\sigma_{x,y}$ la covariance de la série statistique $(x_i, y_i)_{1\leqslant i\leqslant n}$, c'est-à-dire :

$$\sigma_{x,y} = \frac{1}{n}\sum_{i=1}^{n} x_i y_i - \overline{x}\,\overline{y}$$

1. Prouver que $f$ admet un unique point critique $(m, q)$ sur $\mathbb{R}^2$ et le déterminer.
2. (a) Démontrer que :

$$\forall (a,b) \in \mathbb{R}^2,\ f(a,b) = n\left[(b - \overline{y} + a\overline{x})^2 + \left(a\sigma_x - \frac{\sigma_{x,y}}{\sigma_x}\right)^2 + \frac{1}{\sigma_x^2}\left(\sigma_x^2\sigma_y^2 - \sigma_{x,y}^2\right)\right]$$

   (b) En déduire que $f$ admet un minimum global sur $\mathbb{R}^2$, atteint uniquement en $(m, q)$. La droite d'équation $y = mx + q$ est appelée droite des moindres carrés.

**Théorème 27.9**

Soit $((x_i, y_j, n_{i,j}))_{\substack{1\leqslant i\leqslant p\\ 1\leqslant j\leqslant q}}$ une série statistique double telle que les écarts-types des séries marginales en $x$ et en $y$ soient non nuls.
La fonction $f : (a,b) \mapsto \sum\limits_{i=1}^{n} [y_i - (ax_i + b)]^2$ admet un minimum sur $\mathbb{R}^2$ et celui-ci est atteint en un unique couple $(a_0, b_0)$, qui vérifie :

$$a_0 = \frac{\operatorname{Cov}(x,y)}{\mathbb{V}(x)} \quad \text{et} \quad b_0 = \overline{y} - a_0\overline{x}$$

La droite d'équation $y = a_0x + b_0$ par par le point moyen du nuage et constitue donc un ajustement affine de la série, appelé **droite des moindres carrés** ou droite de **droite de régression** de $y$ en $x$.

**Exemple 27.4** Reprenons la première série proposée dans l'exemple 25.11.
Rappelons que la commande `np.cov(x,y,bias=True)` renvoie la matrice :

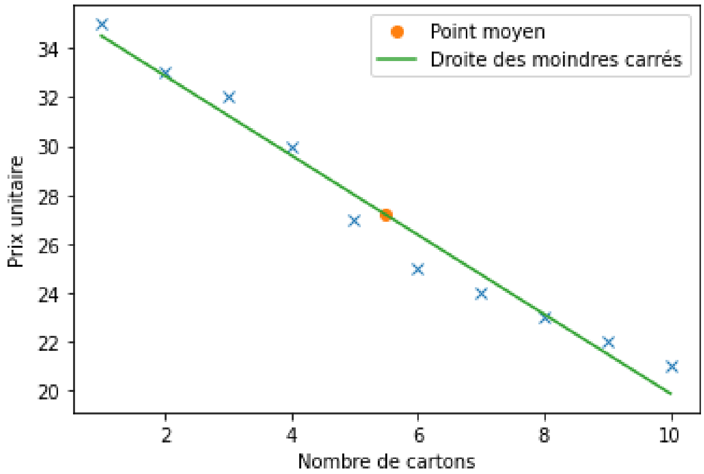
$$\begin{pmatrix} \mathbb{V}(x) & \operatorname{Cov}(x,y) \\ \operatorname{Cov}(x,y) & \mathbb{V}(y) \end{pmatrix}$$

On peut utiliser Python pour déterminer l'équation de la droite des moindres carrés. Ainsi, en exécutant le programme :

```
import numpy as np
import matplotlib.pyplot as plt
fig,ax=plt.subplots()
ax.set_xlabel('Nombre de cartons')
ax.set_ylabel("Prix unitaire")
x=np.arange(1,11)
y=np.array([35,33,32,30,27,25,24,23,22,21])
moy_x=np.mean(x)
moy_y=np.mean(y)
C=np.cov(x,y,bias=True)
a=C[0,1]/C[0,0]
b=np.mean(y)-a*np.mean(x)
plt.plot(x,y,'x')
plt.plot(moy_x,moy_y,'o',label="Point moyen")
plt.plot(x,a*x+b,label='Droite des moindres carrés')
plt.legend()
```

© www.stephanepreteseille.com. Diffusion et reproduction interdites, même partiellement

on obtient la représentation graphique suivante :




# E. Correction des exercices

## Correction de l'exercice 27-1

On peut remarquer que $P$ est un polynôme toujours positif. De plus on a :

$$\begin{aligned}\forall t \in \mathbb{R},\ P(t) &= \frac{1}{n}\sum_{i=1}^{p}\sum_{j=1}^{q} n_{i,j}\left[t^2\,(x_i-\overline{x})^2 - 2t\,(x_i-\overline{x})\,(y_j-\overline{y}) + (y_j-\overline{y})^2\right] \\ &= \frac{t^2}{n}\sum_{i=1}^{p} n_{i,\cdot}\,(x_i-\overline{x})^2 - \frac{2t}{n}\sum_{i=1}^{p}\sum_{j=1}^{q} n_{i,j}\,(x_i-\overline{x})\,(y_j-\overline{y}) + \frac{1}{n}\sum_{j=1}^{q} n_{\cdot,j}\,(y_j-\overline{y})^2 \\ &= t^2\,\mathbb{V}(x) - 2t\,\mathrm{Cov}(x,y) + \mathbb{V}(y)\end{aligned}$$

Ainsi $P$ est un polynôme de degré 2, toujours positif, donc son discriminant est négatif ou nul, c'est-à-dire :

$$[2\,\mathrm{Cov}(x,y)]^2 - 4\,\mathbb{V}(x)\,\mathbb{V}(y) \leqslant 0$$

d'où :

$$[\mathrm{Cov}(x,y)]^2 \leqslant \mathbb{V}(x)\,\mathbb{V}(y)$$

On en déduit, comme $\mathbb{V}(x)\,\mathbb{V}(y) \neq 0$ et comme la fonction $t \mapsto \sqrt{t}$ est croissante sur $\mathbb{R}_+$ :

$$|\sigma(x,y)| \leqslant 1$$

Par ailleurs, on a également :

$$\begin{aligned}|\sigma(x,y)| = 1 &\iff [\mathrm{Cov}(x,y)]^2 = \mathbb{V}(x)\,\mathbb{V}(y) \\ &\iff [2\,\mathrm{Cov}(x,y)]^2 - 4\,\mathbb{V}(x)\,\mathbb{V}(y) = 0 \\ &\iff \exists\, a \in \mathbb{R},\ P(r) = 0 \\ &\iff \exists\, a \in \mathbb{R},\ \frac{1}{n}\sum_{i=1}^{p}\sum_{j=1}^{q} n_{i,j}\,(a\,(x_i-\overline{x}) - (y_j-\overline{y}))^2 = 0\end{aligned}$$

donc, comme tous les termes de la somme sont positifs ou nuls :

$$\begin{aligned}|\sigma(x,y)| = 1 &\iff \exists\, a \in \mathbb{R},\ \forall (i,j) \in [\![1,p]\!] \times [\![1,q]\!],\ (a\,(x_i-\overline{x}) - (y_j-\overline{y}))^2 = 0 \\ &\iff \exists\, a \in \mathbb{R},\ \forall (i,j) \in [\![1,p]\!] \times [\![1,q]\!],\ (y_j-\overline{y}) = a\,(x_i-\overline{x}) \\ &\iff \exists\, a \in \mathbb{R},\ \forall (i,j) \in [\![1,p]\!] \times [\![1,q]\!],\ y_j = a x_i - a\overline{x} + \overline{y}\end{aligned}$$

© www.stephanepreteseille.com

© www.stephanepreteseille.com. Diffusion et reproduction interdites, même partiellement

## Correction de l'exercice 27-2

1. $f$ est une fonction polynôme de classe $\mathcal{C}^1$ sur $\mathbb{R}^2$ et on a :

$$\begin{aligned}\forall (a,b)\in\mathbb{R}^2,\ \partial_1 f(a,b) &= \sum_{i=1}^n -2x_i\left[y_i-(ax_i+b)\right]\\ &= 2a\sum_{i=1}^n x_i^2+2b\sum_{i=1}^n x_i-2\sum_{i=1}^n x_iy_i\\ &= 2\left(a\sum_{i=1}^n x_i^2+bn\overline{x}-\sum_{i=1}^n x_iy_i\right)\end{aligned}$$

et :

$$\begin{aligned}\forall (a,b)\in\mathbb{R}^2,\ \partial_2 f(a,b) &= \sum_{i=1}^n 2\left[y_i-(ax_i+b)\right]\\ &= -2a\sum_{i=1}^n x_i-2nb+2\sum_{i=1}^n y_i\\ &= -2n\left(a\overline{x}+b-\overline{y}\right)\end{aligned}$$

Pour tout $(a,b)\in\mathbb{R}^2$, on a donc :

$$\begin{aligned}\nabla(f)(a,b)=0 &\iff \begin{cases} a\displaystyle\sum_{i=1}^n x_i^2+bn\overline{x}-\sum_{i=1}^n x_iy_i=0\\ a\overline{x}+b-\overline{y}=0\end{cases}\\ &\iff \begin{cases} a\left(\displaystyle\sum_{i=1}^n x_i^2-n\overline{x}^2\right)+n\overline{x}\,\overline{y}-\sum_{i=1}^n x_iy_i=0\\ b=\overline{y}-a\overline{x}\end{cases}\\ &\iff \begin{cases} na\sigma_x^2=n\,\sigma_{x,y}\\ b=\overline{y}-a\overline{x}\end{cases}\end{aligned}$$

et finalement, comme $\sigma_x\neq 0$ (car $x_1,\dots,x_n$ ne sont pas tous égaux) :

$$\nabla(f)(a,b)=0 \iff \begin{cases} a=\dfrac{\sigma_{x,y}}{\sigma_x^2}\\ b=\overline{y}-\overline{x}\,\dfrac{\sigma_{x,y}}{\sigma_x^2}\end{cases}$$

ce qui nous permet de conclure :

$$\boxed{(m,q)=\left(\frac{\sigma_{x,y}}{\sigma_x^2},\overline{y}-\overline{x}\,\frac{\sigma_{x,y}}{\sigma_x^2}\right)\text{ est l'unique point critique de } f}$$

2. (a) Soit $(a,b)\in\mathbb{R}^2$. On a :

$$\begin{aligned}f(a,b) &= \sum_{i=1}^n\left[y_i-(ax_i+b)\right]^2\\ &= \sum_{i=1}^n\left(y_i^2+a^2x_i^2+b^2-2ax_iy_i-2by_i+2abx_i\right)\\ &= \sum_{i=1}^n y_i^2+a^2\sum_{i=1}^n x_i^2+nb^2-2a\sum_{i=1}^n x_iy_i-2b\sum_{i=1}^n y_i+2ab\sum_{i=1}^n x_i\\ &= n\left[(\sigma_y^2+\overline{y}^2)+a^2\left(\sigma_x^2+\overline{x}^2\right)+b^2-2a\left(\sigma_{x,y}+\overline{x}\,\overline{y}\right)-2b\overline{y}+2ab\overline{x}\right]\\ &= n\left[(b^2+\overline{y}^2+a^2\overline{x}^2-2b\overline{y}+2ab\overline{x}-2a\overline{x}\,\overline{y})+\left(a^2\sigma_x^2-2a\sigma_{x,y}+\frac{\sigma_{x,y}^2}{\sigma_x^2}\right)+\left(\sigma_y^2-\frac{\sigma_{x,y}^2}{\sigma_x^2}\right)\right]\end{aligned}$$

© www.stephanepreteseille.com. Diffusion et reproduction interdites, même partiellement

et finalement :

$$\boxed{\forall (a,b)\in\mathbb{R}^2,\ f(a,b)=n\left[(b-\overline{y}+a\overline{x})^2+\left(a\sigma_x-\frac{\sigma_{x,y}}{\sigma_x}\right)^2+\frac{1}{\sigma_x^2}\left(\sigma_x^2\sigma_y^2-\sigma_{x,y}^2\right)\right]}$$

**N.B.** Il était évidemment aussi possible (et sans doute plus sage) de développer le résultat proposé et de le comparer avec la première expression développée de $f(a,b)$.

(b) Comme $f$ est de classe $\mathcal{C}^1$ sur $\mathbb{R}^2$ et comme $\mathbb{R}^2$ est ouvert, $f$ ne peut admettre d'extremum qu'en $(m,q)$. De plus, on a, d'après le calcul précédent :

$$f(m,q)=\frac{1}{\sigma_x^2}\left(\sigma_x^2\sigma_y^2-\sigma_{x,y}^2\right)$$

et donc :

$$\begin{aligned}\forall (a,b)\in\mathbb{R}^2,\ f(a,b)-f(m,q)&=n\left[(b-\overline{y}+a\overline{x})^2+\left(a\sigma_x-\frac{\sigma_{x,y}}{\sigma_x}\right)^2\right]\\&\geqslant 0\end{aligned}$$

ce qui nous permet de conclure :

$$\boxed{f \text{ admet un minimum global sur } \mathbb{R}^2\text{, atteint uniquement en } (m,q)}$$

www.stephanepreteseille.com ©

© www.stephanepreteseille.com. Diffusion et reproduction interdites, même partiellement

# Sommaire



© www.stephanepreteseille.com