Société de Calcul Mathématique SA
Outils d'aide à la décision
depuis 1995

CLQ_2019_03

Exposé de Bernard Beauzamy

I. Comparaison de deux ensembles de données

pour le même site : Lille – Lesquin

sources :

Infoclimat
Association Loi 1901 (droit français)
https://www.infoclimat.fr/
le site n'est soutenu que par les dons et adhésions

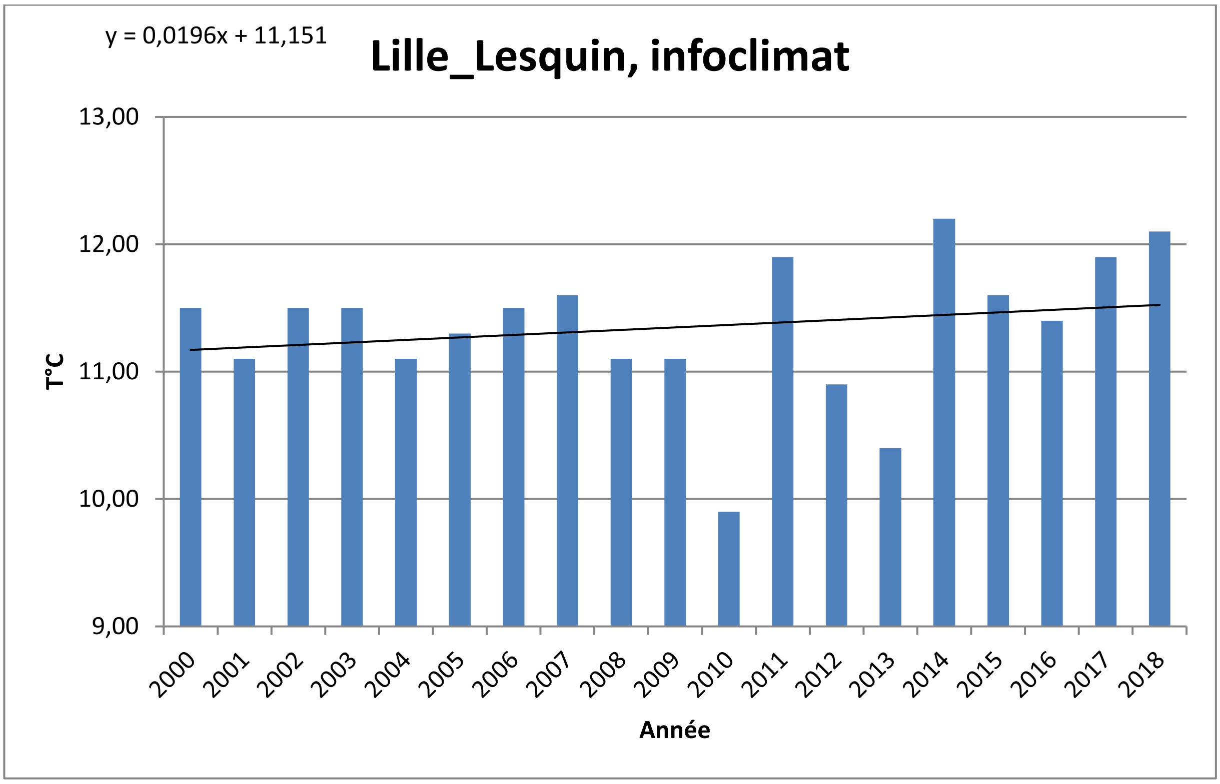
et ECAD
European Climate Assessment & Dataset project https://www.ecad.eu/

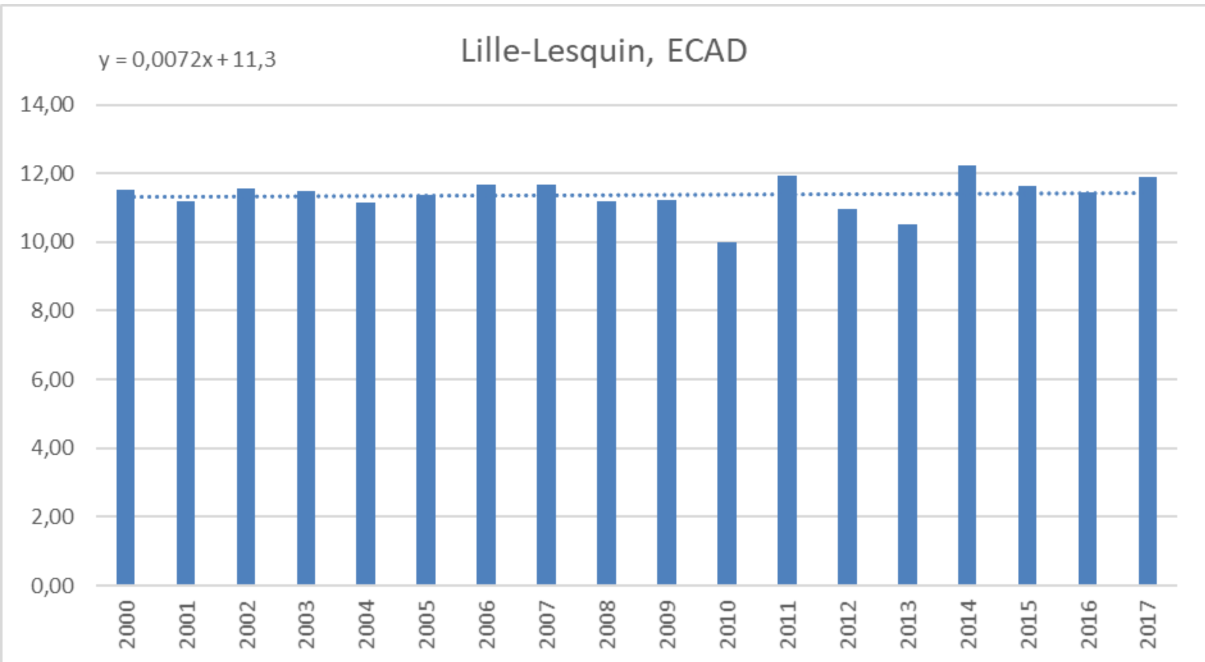
année	temp infoclimat	temp ecad
2000	11,50	11,52
2001	11,10	11,17
2002	11,50	11,56
2003	11,50	11,50
2004	11,10	11,15
2005	11,30	11,39
2006	11,50	11,67
2007	11,60	11,69
2008	11,10	11,17
2009	11,10	11,21
2010	9,90	9,98
2011	11,90	11,94
2012	10,90	10,97
2013	10,40	10,50
2014	12,20	12,22
2015	11,60	11,65
2016	11,40	11,44
2017	11,90	11,90
2018	12,10	???
moyenne	11,31	11,37

On constate que InfoClimat tronque à la première décimale (sans arrondi), mais il y a des différences, voir 2006.

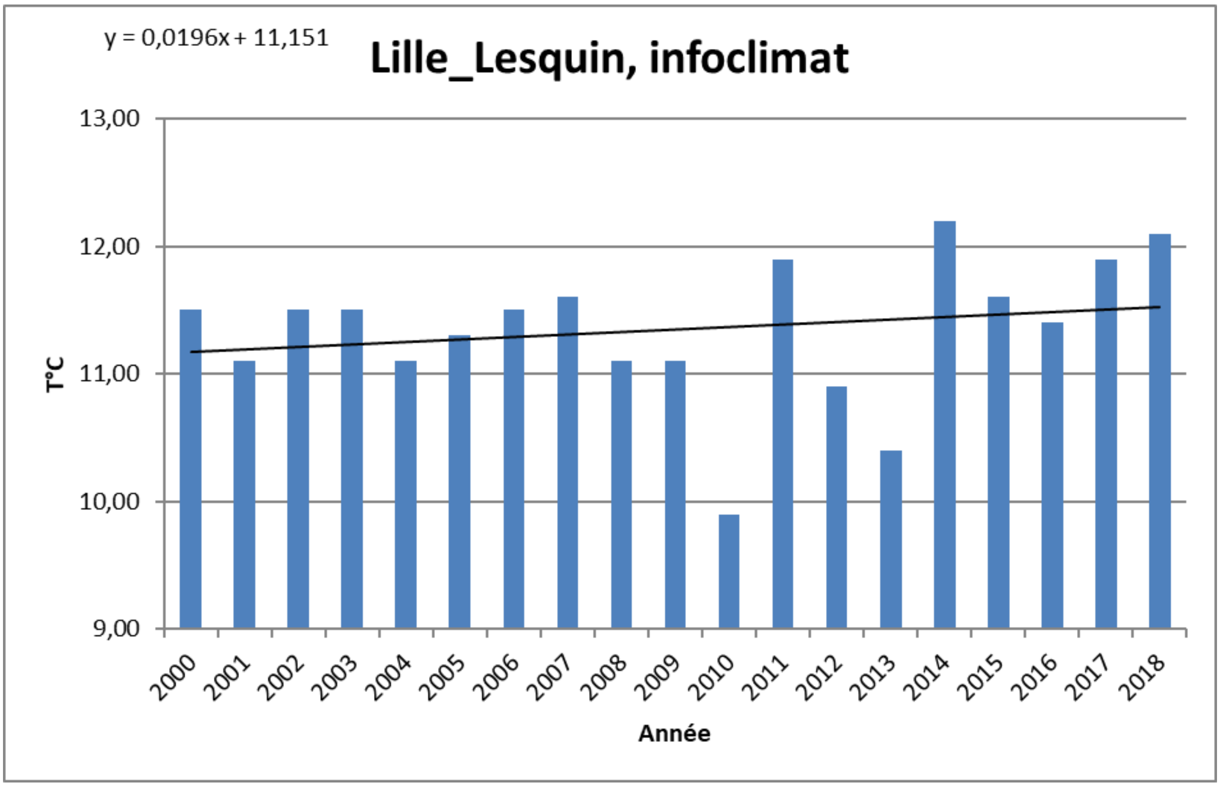
Dans le cas de ECAD, les températures sont journalières ; les moyennes annuelles sont calculées par nous. Il s'agit de la température TG (mean of daily mean temperature : température moyenne par jour, moyennée sur l'année).

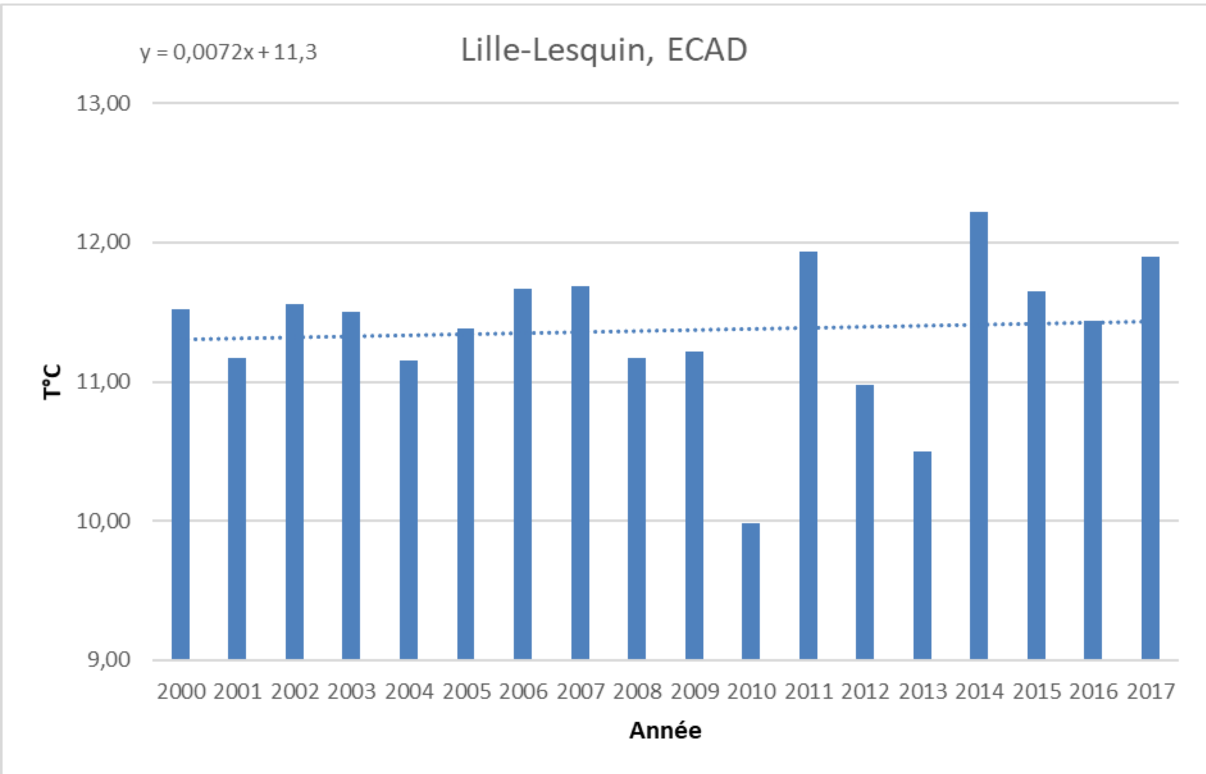
Déterminons la droite de tendance dans chaque cas :





On constate avec amusement que ces droites ont des aspects fondamentalement différents. Comme fait remarquer Zoe Louyot, la différence provient largement du fait que l'échelle en ordonnées n'est pas la même. Voici les deux diagrammes avec la même échelle :





II. L'effet du hasard

Avant de conclure qu'il y a une tendance au réchauffement, ou au refroidissement, sur 100 ans, voyons l'effet du hasard : comment se comporte la droite de tendance si les températures sont générées uniquement par une loi uniforme.

Commençons par rappeler comment la droite de tendance est définie :

1. Rappels de définition

Etant donnés des points (x_k, y_k) $k = 1,...,N$, la droite de tendance minimise la quantité :

$$D = \sum_{k=1}^{N} \left(y_k - (ax_k + b)\right)^2$$

Les deux équations $\frac{\partial D}{\partial a} = 0, \frac{\partial D}{\partial b} = 0,$ se traduisent par :

$$\begin{aligned} &\sum_{k=1}^{N} x_k\left(y_k - (ax_k + b)\right) = 0 \\ &\sum_{k=1}^{N} y_k - (ax_k + b) = 0 \end{aligned} \qquad (1)$$

En notant $m_x = \frac{1}{N}\sum_{k=1}^{N} x_k, m_y = \frac{1}{N}\sum_{k=1}^{N} y_k$, on déduit de (1b) :

$$b = m_y - am_x$$

et en reportant dans (1a) :

$$\sum_{k=1}^{N} x_k\left(y_k - ax_k - m_y + am_x\right) = 0$$

ou :

$$\sum_{k=1}^{N} x_k y_k - ax_k^2 - m_y x_k + am_x x_k = 0$$

et donc :

$$a = \frac{\operatorname{cov}(X,Y)}{\operatorname{var}(X)}$$

en notant :

$$\operatorname{cov}(X,Y)=\frac{1}{N}\sum_{k=1}^{N}\ x_k y_k-m_x m_y$$

$$\operatorname{var}(X)=\frac{1}{N}\sum_{k=1}^{N}\ x_k^2-m_x^2$$

2. Cas d'une série temporelle

Dans le cas d'une série temporelle, où les abscisses sont en années, avec l'origine des temps en l'an T_0 :

$$x_k=T_0+k,\ \ k=1,\dots,N$$

$$m_x=\frac{1}{N}\sum_{i=1}^{N}(T_0+k)=T_0+\frac{N+1}{2}$$

$$\operatorname{var}(X)=\frac{N^2-1}{12}$$

$$\operatorname{cov}(X,Y)=\frac{1}{N}\sum_{k=1}^{N}\ (T_0+k)\,y_k-\left(T_0+\frac{N+1}{2}\right)m_y=\frac{1}{N}\sum_{k=1}^{N}\ ky_k-\frac{N+1}{2}m_y$$

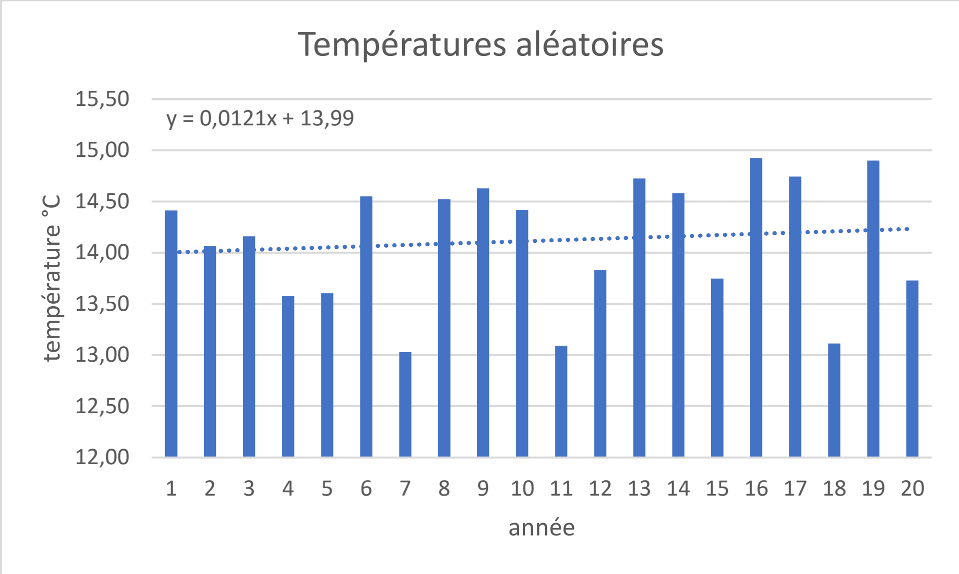
et donc le coefficient directeur de la droite de tendance vaut :

$$a=\frac{\dfrac{1}{N}\displaystyle\sum_{k=1}^{N}\ ky_k-\dfrac{N+1}{2}m_y}{\dfrac{N^2-1}{12}}$$

Il est indépendant de T_0.

3. Cas des valeurs d'une température

Voyons maintenant le cas où les y_k sont les valeurs d'une température, issues d'une loi uniforme entre 13°C et 15°C. Voici un exemple :



On a ici, avec $\vartheta_0 = 13°C$

$$y_k = \vartheta_0 + 2X_k$$

où les X_k sont des variables aléatoires indépendantes, suivant une loi uniforme sur $[0,1]$. Réécrivons la formule du §2 :

$$a = \frac{\frac{1}{N}\sum_{k=1}^{N} ky_k - \frac{N+1}{2}m_y}{\frac{N^2-1}{12}} \tag{1}$$

On a ici :

$$m_y = \frac{1}{N}\sum_{k=1}^{N} y_k = \frac{1}{N}\sum_{k=1}^{N} E\left(\vartheta_0 + 2X_k\right) = \vartheta_0 + 2\frac{1}{N}\sum_{k=1}^{N} E\left(X_k\right) = \vartheta_0 + 1$$

Par ailleurs :

$$\sum_{k=1}^{N} ky_k = \sum_{k=1}^{N} k\left(\vartheta_0 + 2X_k\right) = \frac{N\left(N+1\right)}{2}\vartheta_0 + 2\sum_{k=1}^{N} kX_k$$

et donc :

$$\frac{1}{N}\sum_{k=1}^{N} ky_k = \frac{N+1}{2}\vartheta_0 + \frac{2}{N}\sum_{k=1}^{N} kX_k$$

On obtient :

$$a=\frac{12}{N^2-1}\left(\frac{N+1}{2}\vartheta_0+\frac{2}{N}\sum_{k=1}^{N}kX_k-\frac{N+1}{2}(\vartheta_0+1)\right)$$
$$=\frac{12}{N^2-1}\left(\frac{2}{N}\sum_{k=1}^{N}kX_k-\frac{N+1}{2}\right)=\frac{24}{N\left(N^2-1\right)}\sum_{k=1}^{N}kX_k-\frac{6}{N-1}$$

et finalement :

$$a=\frac{24}{N\left(N^2-1\right)}\sum_{k=1}^{N}kX_k-\frac{6}{N-1} \tag{2}$$

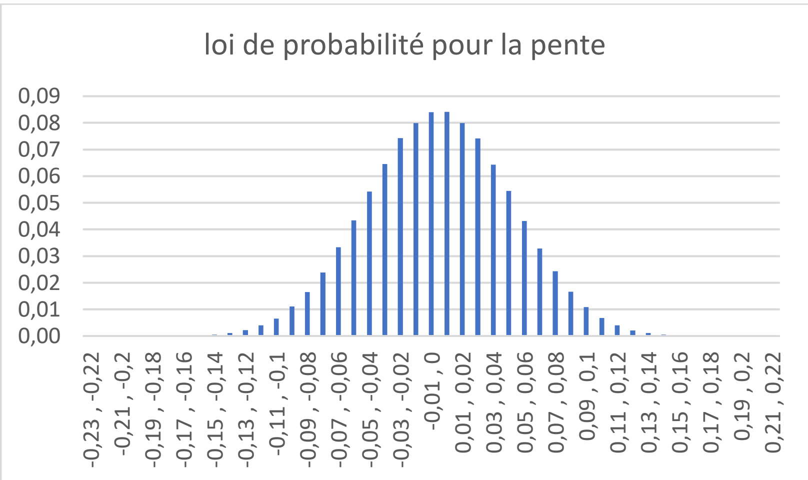
L'espérance de a est nulle, évidemment. La variance vaut (puisque $\mathrm{var}\left(X_k\right)=\frac{1}{12}$) :

$$\mathrm{var}(a)=\mathrm{var}\left(\frac{24}{N\left(N^2-1\right)}\sum_{k=1}^{N}kX_k-\frac{6}{N-1}\right)=\mathrm{var}\left(\frac{24}{N\left(N^2-1\right)}\sum_{k=1}^{N}kX_k\right)=\left(\frac{24}{N\left(N^2-1\right)}\right)^2\mathrm{var}\left(\sum_{k=1}^{N}kX_k\right)$$
$$=\left(\frac{24}{N\left(N^2-1\right)}\right)^2\sum_{k=1}^{N}\mathrm{var}\left(kX_k\right)=\left(\frac{24}{N\left(N^2-1\right)}\right)^2\sum_{k=1}^{N}k^2\,\mathrm{var}\left(X_k\right)$$
$$=\left(\frac{24}{N\left(N^2-1\right)}\right)^2\frac{1}{12}\sum_{k=1}^{N}k^2=\frac{48}{N^2\left(N^2-1\right)^2}\frac{1}{6}N(N+1)(2N+1)=\frac{8(2N+1)}{N(N+1)(N-1)^2}$$

et ceci est équivalent à $\frac{16}{N^3}$, et donc $\to 0$ lorsque $N\to+\infty$.

4. Cas particulier

Cas $N=20$ (utilisé pour les données de température). On obtient la loi de probabilité suivante :



La probabilité que $a\geq 0,03$ (ce qui correspond à un accroissement moyen de 3°C sur 100 ans) est $\geq 0,26$. Bien sûr, la probabilité que $a\leq -0,03$ est la même. En définitive, la probabilité que le pronostic sur 100 ans soit "gentiment" entre -3°C et 3°C est $p \approx 0,476$. Autrement dit, un simple tirage aléatoire de températures, selon une loi uniforme, donne plus d'une fois sur deux un pronostic inquiétant sur 100 ans.

Dans le cas $N = 18$, la droite de tendance peut varier davantage, et la probabilité que le pronostic sur 100 ans soit "gentiment" entre -3°C et 3°C est $p \approx 0,412$.

Autrement dit, dans plus d'un cas sur deux, le fait que le pronostic à 100 ans soit supérieur à $3°C$, ou inférieur à $-3°C$, s'explique simplement par le fait du hasard : une loi uniforme sur les températures produit cet effet.