

Statistiques descriptives des concentrations horaires des contaminants mesurées de façon continue par le Réseau de surveillance de la qualité de l'air du Québec

Les statistiques descriptives permettent de résumer et d'interpréter plus facilement des ensembles de données complexes.

Toutes les statistiques descriptives des données horaires mesurées de façon continue par le Réseau de surveillance de la qualité de l'air du Québec (RSQAQ) ont été calculées à partir des fichiers disponibles sur Données Québec dans la section RSQAQ - Données horaires continues (<https://www.donneesquebec.ca/recherche/dataset/rsqaq-donnees-horaires-continues>).

Ce document présente une explication détaillée des calculs effectués pour déterminer les valeurs des statistiques descriptives du fichier publié.

Moyenne arithmétique

Le calcul de la moyenne arithmétique (\bar{x}) a été effectué selon la formule suivante :

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

où x_i représente chaque concentration individuelle dans l'ensemble de données et n est le nombre total de concentrations.

Écart-type

L'écart-type (s) est une mesure de la dispersion des concentrations d'un échantillon statistique ou d'une distribution de probabilités. Le calcul de l'écart-type a été effectué selon la formule suivante :

$$s = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n - 1}}$$

où x_i représente chaque concentration individuelle dans l'ensemble de données, \bar{x} est la moyenne arithmétique des concentrations, et n est le nombre total de concentrations.

Minimum

Le minimum représente la concentration minimale d'un ensemble de données. Dans ce cas, il s'agit du minimum des concentrations pour chaque ensemble de stations, année et contaminant.

Centiles

Pour le calcul des centiles, l'approche de Rick Wicklin (2017) a été utilisée. Cette approche correspond au type 2 de la formule des centiles du « package stats » du logiciel R (R Core Team, 2022). Elle calcule une interpolation linéaire entre deux points de données adjacents, comme expliqué ci-dessous :

1. Calculer la position : On commence par déterminer la position j du centile dans l'échantillon en utilisant la formule $j = \lfloor Np \rfloor$, où N est la taille de l'échantillon et p est le centile souhaité.
2. Déterminer le poids : On calcule ensuite le poids λ avec $\lambda = Np - j$.
3. Interpoler : On utilise ces valeurs pour interpoler entre les points de données $x_{\lfloor j \rfloor}$ et $x_{\lfloor j+1 \rfloor}$ avec la formule suivante : $q = (1 - \lambda)x_{\lfloor j \rfloor} + \lambda x_{\lfloor j+1 \rfloor}$.

Maximum

Le maximum représente la concentration maximale d'un ensemble de données, dans ce cas, le maximum des concentrations pour chaque ensemble de stations, année et contaminant.

Références

R Core Team (2022). R : A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

Wicklin, R. (2017). *Sample quantiles: A comparison of 9 definitions*. The DO Loop. SAS Blogs