

Statistiques descriptives des concentrations de métaux et de particules mesurées de façon séquentielle 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 mesurées de façon séquentielle 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 de métaux et de particules séquentielles » (<https://www.donneesquebec.ca/recherche/dataset/rsqaq-donnees-sequentielles>).

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

Traitement des valeurs sous la limite de détection

Les résultats inférieurs à la limite de détection (LD) ont été remplacés par la moitié de cette limite (LD/2) pour le calcul des statistiques descriptives.

Nombre de valeurs totales (n)

Nombre total de valeurs pour chaque combinaison de station, d'année et de contaminant.

Nombre de valeurs supérieures à la limite de détection (n_plus_grande_ld)

Nombre de valeurs supérieures à la limite de détection (LD).

Pourcentage de détection (detection)

Pourcentage de valeurs supérieures à la limite de détection (LD), calculé selon la formule suivante :

$$\text{Détection (\%)} = \frac{n > LD}{n} \times 100$$

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 valeur individuelle dans l'ensemble de données et n est le nombre total de valeurs.

Ecart-type

L'écart-type (s) est une mesure de la dispersion des valeurs 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 valeur individuelle dans l'ensemble de données, \bar{x} est la moyenne arithmétique des valeurs, et n est le nombre total de valeurs.

Minimum

Le minimum représente la valeur minimale d'un ensemble de données. Dans ce cas, il s'agit du minimum des concentrations pour chaque ensemble de station, 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 quantile du « paquet stats » du logiciel R (R Core Team, 2022). Elle calcule une interpolation linéaire entre deux points de données adjacents, selon les formules ci-dessous :

1. Calculer la position : la position j du quantile dans l'échantillon est déterminée en utilisant la formule $j = \lfloor Np \rfloor$ où N est la taille de l'échantillon et p est le quantile souhaité.
2. Déterminer le poids : le poids λ est calculé avec la formule $\lambda = Np - j$.
3. Interpoler : ces valeurs sont utilisées pour interpoler la valeur entre les points de données $x_{[j]}$ et $x_{[j+1]}$ avec la formule suivante : $q = (1 - \lambda)x_{[j]} + \lambda x_{[j+1]}$.

Maximum

Le maximum représente la valeur maximale d'un ensemble de données. Dans ce cas, il s'agit du maximum des concentrations pour chaque ensemble de station, 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