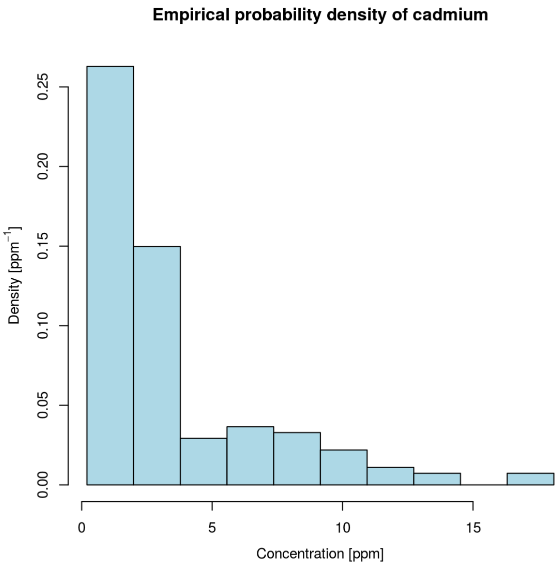
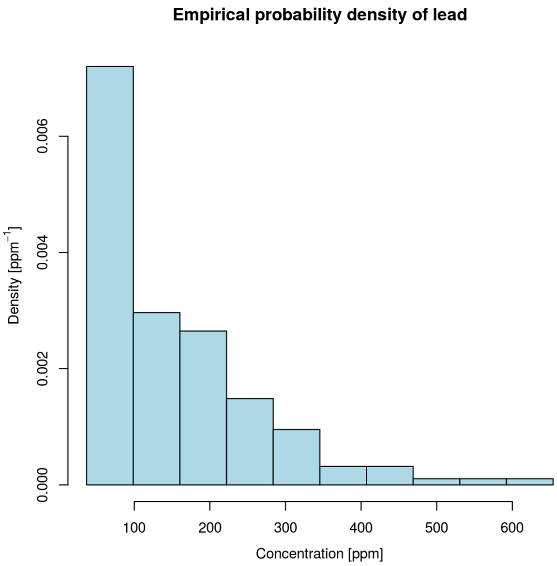
Exercise 3 - Variogram fitting - Solutions

Office hours: Friday 09:00-12:00

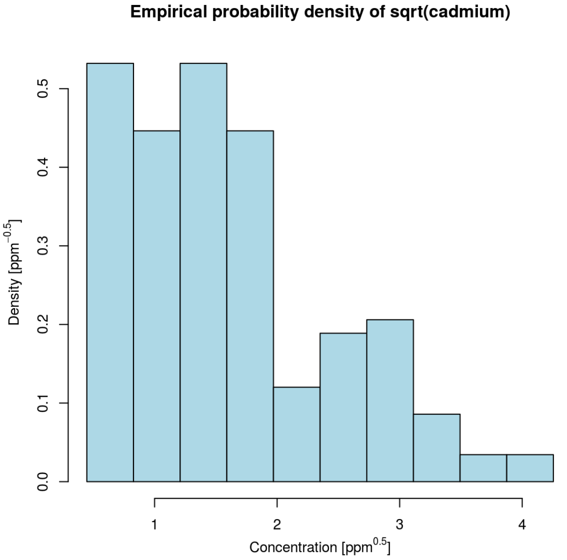
A. Berne alexis.berne@epfl.ch
G. Ghiggi gionata.ghiggi@epfl.ch
M. Guidicelli matteo.guidicelli@epfl.ch

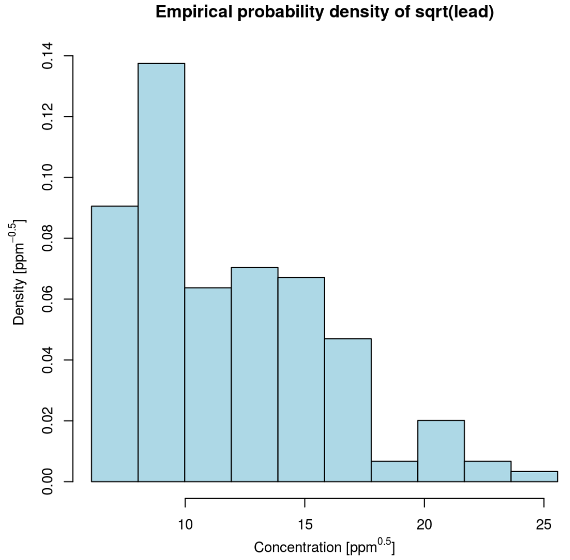
1. *Plot and show the empirical probability density function (pdf) of the cadmium and lead concentrations.*

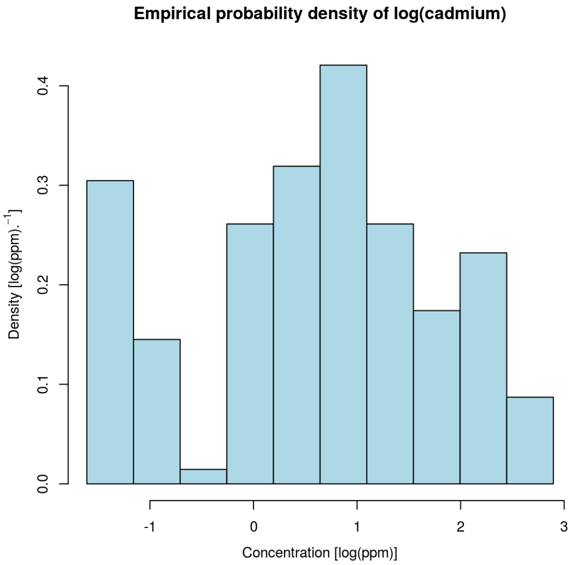


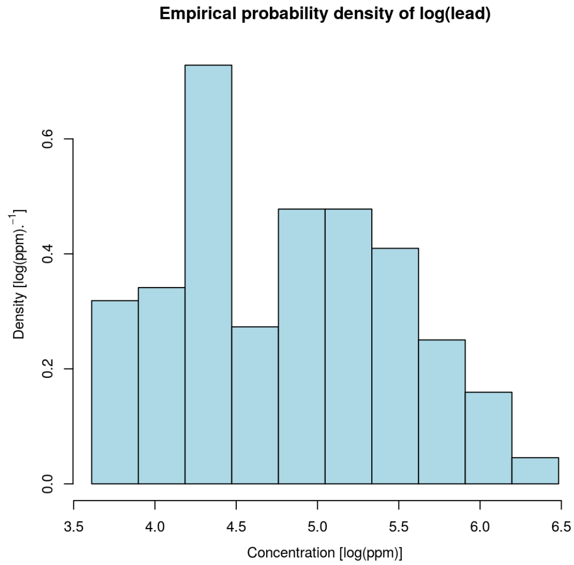


2. *Apply a square root transform on the values of cadmium and lead. Show the empirical probability density function of the transformed concentration values. Do the same using a log transform.*







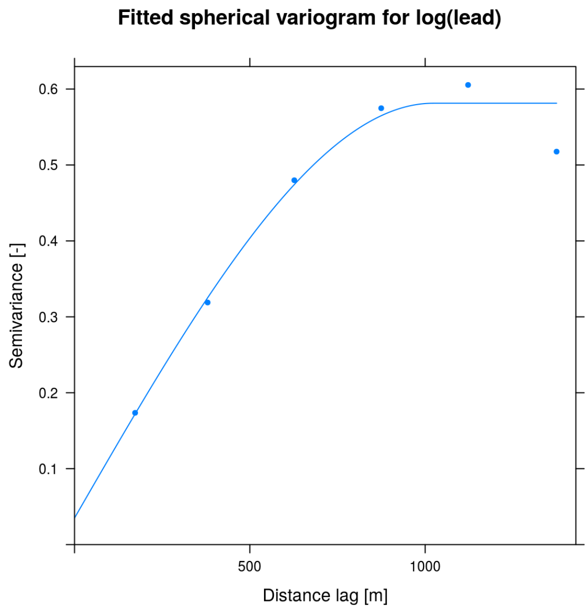


3. *Apply the Shapiro-Wilk test to assess which transformation (sqrt or log) provides the best results, i.e., the most Gaussian like distributions.*

   The Shapiro-Wilk test shows that the log transform clearly provides better results for both variables.

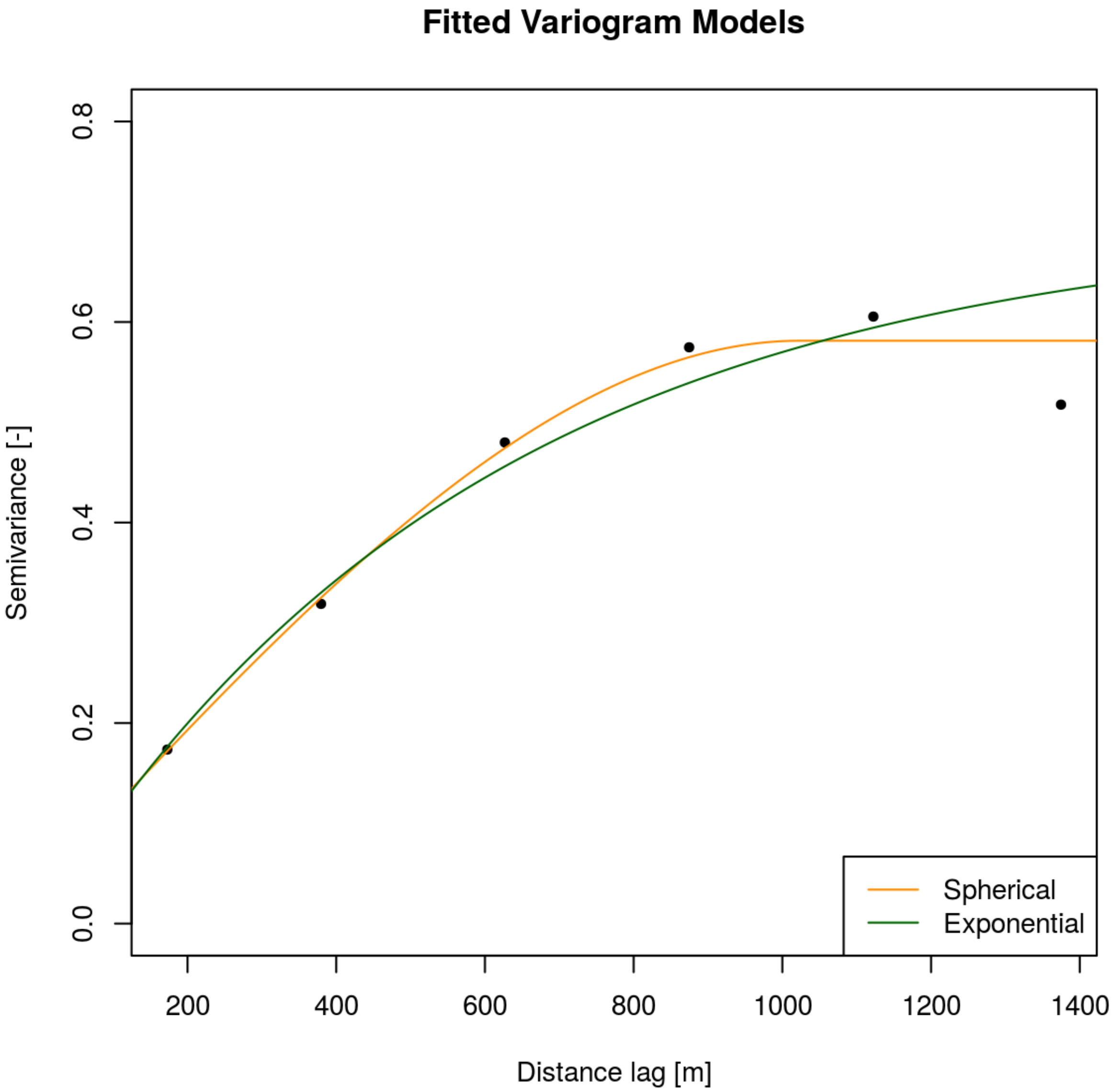
	cadmium	lead
linear	0.786	0.839
sqrt	0.928	0.929
log	0.946	0.970

4. *Compute the isotropic sample variogram of log(cadmium) and fit a spherical variogram model on it. Show the sample variogram and the fitted model on the same graph. What are the fitted values of the nugget, the sill and the range?*



The fitted nugget is 0.586 $\log(\text{ppm})^2$, the sill is 1.913 $\log(\text{ppm})^2$ and the range is 1235.2 m.

5. *Compute the isotropic sample variogram of log(lead). Fit both a spherical and an exponential variogram model on it. Which fit is the best?*



The spherical variogram model provides the best fit. It has a weighted sum of squared errors of $2.551 \cdot 10^{-6}$ against $9.880 \cdot 10^{-6}$ for the exponential model. The spherical model also better captures the range and the sill of the sample variogram.

6. Show that $\mathrm{Var}\left[\sum_{i=1}^{n} \lambda_i X_i\right] = \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \lambda_j \mathrm{Cov}[X_i, X_j]$

$$
\begin{aligned}
\mathrm{Var}\left[\sum_{i=1}^{n}\lambda_i X_i\right] &= \mathrm{E}\left[(\sum_{i=1}^{n}\lambda_i X_i)^2\right] - \left(\mathrm{E}[\sum_{i=1}^{n}\lambda_i X_i]\right)^2 \\
&= \mathrm{E}\left[\sum_{i=1}^{n}\sum_{j=1}^{n}\lambda_i\lambda_j X_i X_j\right] - \left(\sum_{i=1}^{n}\lambda_i \mathrm{E}[X_i]\right)^2 \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n}\lambda_i\lambda_j \mathrm{E}[X_i X_j] - \sum_{i=1}^{n}\sum_{j=1}^{n}\lambda_i\lambda_j \mathrm{E}[X_i]\mathrm{E}[X_j] \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n}\lambda_i\lambda_j \left(\mathrm{E}[X_i X_j] - \mathrm{E}[X_i]\mathrm{E}[X_j]\right) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n}\lambda_i\lambda_j \mathrm{Cov}[X_i, X_j]
\end{aligned}
$$