Suppose we have two vectors of data, \( \mathbf{X} = (X_1, \ldots, X_n)^T \) and \( \mathbf{Y} = (Y_1, \ldots, Y_n)^T \).

Now we think about the regression problem (linear regression) of \( \mathbf{Y} \) on \( \mathbf{X} \).

Our model is \( \mathbf{Y} = a \mathbf{X} + b + \mathbf{e} \), where \( a \) and \( b \) are parameters to be estimated and we assume \( \mathbf{e} \sim N(0, \sigma^2 I) \).

Then we find \( a \) and \( b \) that minimize the sum of squares
\[
\sum_{i=1}^{n} (Y_i - a - b \cdot X_i)^2
\]
This is called the Least Squares (LS) method.
Estimating variogram using LS

- Suppose $\gamma(h; \theta)$ is the true variogram and we have empirical variogram values at $n$ distance lag: $\hat{\gamma}(h_i), i = 1, \cdots, n$

- We consider a model $\hat{\gamma}(h) = \gamma(h; \theta) + e(h)$ and we assume $e$ is mean zero.

- The tricky part is the covariance structure of $e$
  - Since we have $n$ “observed” variograms at distance lags $h_i, i = 1, \cdots, n$, we need to consider a covariance matrix of $e$ with dimension $n \times n$.
  - Are $e(h_i)$ and $e(h_j)$ for $i \neq j$ independent?
  - How can we calculate $\text{Cov}\{e(h_i), e(h_j)\}$?

- If we denote the covariance matrix of $e$ by $R(\theta)$, we need to find $\theta$ that minimizes

\[
\left\{\hat{\gamma}(h) - \gamma(h; \theta)\right\}^T R(\theta)^{-1} \left\{\hat{\gamma}(h) - \gamma(h; \theta)\right\}
\]
Simplest thing to do: OLS

- OLS stands for ordinary least squares
- Here we pretend that $e(h_i)$'s are independent and have a constant variance

- We let $R(\theta) = \phi I$
- Then the problem reduces to finding $\theta$ that minimizes

$$\{\hat{\gamma}(h) - \gamma(h; \theta)\}^T \{\hat{\gamma}(h) - \gamma(h; \theta)\}$$
Next simplest thing to do: WLS

- If we do not like the independent and constant variance assumption, we can do a little more sophisticated thing.
- Here we still pretend that $e(h_i)$’s are independent but we replace the diagonal of $R(\theta)$ by
  \[
  \text{Var}\{\hat{\gamma}(h_i)\} \approx 2 \frac{\gamma(h_i)^2}{|N(h_i)|}
  \]
  - $|N(h_i)|$ denotes the number of observations in the bin that corresponds to the lag $h_i$.
- Some papers in the literature show that the estimators from OLS and WLS are more or less equally efficient (or equally bad).
WLS is still not good since it ignores the off diagonals of $R(\theta)$ while there should be clearly dependence between $e(h_i)$ and $e(h_j)$, $i \neq j$

We can work out the expressions for the covariance matrix of $e$, $R(\theta)$

If we denote $T_{ij} = Z(s_i) - Z(s_j)$ and $h_{ij} = |s_i - s_j|$, then we can show that

$$\text{Cov}(T_{ij}^2, T_{kl}^2) = 2\{\gamma(h_{ij}, \theta) + \gamma(h_{jk}, \theta) - \gamma(h_{jl}, \theta) - \gamma(h_{ik}, \theta)\}$$

While clearly GLS should perform better than OLS or WLS, there is a great computational difficulty

In practice, people may use OLS or just use maximum likelihood estimation
Homework 3 (revised)

1. Variogram:
   - Using the data set (data 1) from the course webpage, plot variograms (try variogram cloud, lines that give binned averages, boxplots of variograms, directional variograms, etc)

2. Optimization problem: (due Oct 6)
   - Download the data file (data 2) from the course webpage
   - Fit a function \( y = f(x) = ax^2 + b + \epsilon(x) \) to the data when \( a \) and \( b \) are parameters
   - Try least squares and assume \( \epsilon(x) \sim N(0, \sigma^2) \) for a constant \( \sigma \)
   - Specifically use “nlm” function in R and get numerical estimates of \( a \) and \( b \)
   - Try at least 5 different starting points (initial values)
   - Report your convergence result (parameter estimates, number of iteration taken) for each starting points

Note: this is NOT a team project and this is NOT for statistics major students only. Every registered students should do this homework individually!
Kriging is a name for Best Linear Unbiased Prediction for spatial statistics
Suppose a random field $Z$ has the following representation:

$$Z(s) = m(s)\beta + \epsilon(s), \epsilon(s) \sim N(0, \Sigma)$$

Note that due to the spatial dependence, $\Sigma$ is not a diagonal matrix in general.

Suppose now you observe the random field at locations $s_i$, $i = 1, \cdots, n$.

Now you wish to predict the value of $Z$ at a new location $s_0$.

This is a general problem of Kriging.

There are many different types of Kriging, mainly due to different mean structure.
In the situation of the previous slide, we express our estimator for $Z(s_0)$ as a linear combination of the observations:

$$\hat{Z}(s_0) = \sum_{i=1}^{n} a_i Z(s_i)$$

Now the question is how to estimate $a_i$'s.

We give two constraints:

- Unbiasedness: $E \hat{Z}(s_0) (= m(s_0)\beta) = \sum_{i=1}^{n} a_iEZ(s_i) = \sum_{i=1}^{n} a_i m(s_i)\beta$
- Best: minimize $E\{Z(s_0) - \hat{Z}(s_0)\}^2$
Different types of Kriging

- Simple Kriging: when the mean is known or when the mean is zero
- Ordinary Kriging: when the mean is an unknown constant
- Universal Kriging: for more general $m$
If we denote the Kriging estimator as $\lambda \mathbf{Z}$ (that is, $\lambda$ is a $1 \times n$ vector with $a_i$'s as its elements), we can easily show that

$$
\lambda = \{ K^{-1} - K^{-1} M (M^T K^{-1} M)^{-1} M^T K^{-1} \} k \\
+ K^{-1} M (M^T K^{-1} M)^{-1} m(s_0)
$$

where $M = (m(s_1), \ldots, m(s_n))^T$, $k = \text{Cov}\{ \mathbf{Z}, \mathbf{Z}(s_0) \}$, and $K = \text{Cov}(\mathbf{Z}, \mathbf{Z}^T)$.

The estimator for $\beta$ is given by $
\hat{\beta} = (M^T K^{-1} M)^{-1} M^T K^{-1} \mathbf{Z}
$

Here we are assuming that $M$ is of full rank and $K$ is nonsingular.
Kriging in R

- Each package, fields and geoR, should have many functions that do Kriging, plotting Krigged surface etc (for example, fields has a function called `Krig`)
- For these functions, you have to specify the covariance model (some of the known models are implemented)
- For more complex covariance models, you can code yourself (this is usually what I do)