Geostatistics - Semivariograms and Covariograms

- The topic of these notes is semivariogram and covariance function (or covariogram) analysis and estimation. We will pick and choose the topics we cover and the depth of coverage. The results will be primarily applicable to geostatistical data, although they find use in lattice (regional) processes also. And the functions we will learn about here will be useful models for spatial correlation structures in spatial regression models. Much of this material is from Chapter 4 of Schabenberger and Gotway.

- We have seen examples of the effect of violating the classical assumption of statistical inference: the observations are a simple random sample from some population of interest, i.e. they are independent and identically distributed random variables. Estimation and prediction now require estimation of a covariance structure.

- We have a spatial process
$$\left\{Z(\mathbf{s}) : \mathbf{s} \in D \subset \mathcal{R}^d\right\}$$
and define
$$\gamma^*(\mathbf{s}_i, \mathbf{s}_j) = \frac{1}{2}\text{Var}\left[Z(\mathbf{s}_i) - Z(\mathbf{s}_j)\right]$$
If γ^* is a function of coordinate distance only then we write
$$\gamma^*(\mathbf{s}_i, \mathbf{s}_j) \equiv \gamma(\mathbf{s}_i - \mathbf{s}_j)$$
and call γ the semivariogram. There is confusion between the use of the terms variogram and semivariogram. Most statistical software packages, including those in `R`, that have semivariogram estimation capability actually use functions or procedures named *variogram*.

- Recall that a spatial process
$$\left\{Z(\mathbf{s}) : \mathbf{s} \in D \subset \mathcal{R}^d\right\}$$
is *intrinsically stationary* if
$$\begin{aligned} E(Z(\mathbf{s})) &= \mu, \quad \text{for all} \quad \mathbf{s} \in D \\ \frac{1}{2}\text{Var}\left[Z(\mathbf{s}_i) - Z(\mathbf{s}_j)\right] &= \gamma(\mathbf{s}_i - \mathbf{s}_j), \quad \text{for all} \quad \mathbf{s}_i, \mathbf{s}_j \in D \end{aligned}$$
Under intrinsic stationarity
$$\gamma^*(\mathbf{s}_i, \mathbf{s}_j) = \frac{1}{2}\text{Var}\left[Z(\mathbf{s}_i) - Z(\mathbf{s}_j)\right] = \frac{1}{2}E\left[\left(Z(\mathbf{s}_i) - Z(\mathbf{s}_j)\right)^2\right]$$
If
$$\gamma(\mathbf{s}_i - \mathbf{s}_j) = \gamma^o(\|\mathbf{s}_i - \mathbf{s}_j\|)$$

then the process is said to be isotropic, otherwise it is anisotropic. Generally we will drop the superscript o when discussing isotropic semivariograms. Recall that γ is a valid semivariogram if and only if it is conditionally negative definite:

$$\sum_{i=1}^{k}\sum_{j=1}^{k} a_i a_j \gamma\left(\mathbf{s}_i - \mathbf{s}_j\right) \leq 0$$

for any finite collection of spatial locations $\mathbf{s}_i, i = 1, \cdots, k$ and real numbers $a_i, i = 1, \cdots, k$ such that $\sum_{i=1}^{k} a_i = 0$.

- The covariance function is defined to be

$$\mathrm{Cov}\left(\mathbf{s}_i, \mathbf{s}_j\right) = C\left(\mathbf{s}_i, \mathbf{s}_j\right) = E\left[\left(Z\left(\mathbf{s}_i\right) - \mu(s_i)\right)\left(Z\left(\mathbf{s}_j\right) - \mu(s_j)\right)\right].$$

If C is a function of $\mathbf{s}_i - \mathbf{s}_j$ only, then we write

$$\mathrm{Cov}\left(\mathbf{s}_i, \mathbf{s}_j\right) = C\left(\mathbf{s}_i - \mathbf{s}_j\right).$$

C is often referred to as the covariogram. In time series analysis it is called the autocovariance function.

- Recall that a spatial process is said to be *second order stationary* if

$$\begin{aligned} E\left(Z\left(\mathbf{s}\right)\right) &= \mu, \quad \text{for all} \quad \mathbf{s} \in D \\ \mathrm{Cov}\left[Z\left(\mathbf{s}_i\right) - Z\left(\mathbf{s}_j\right)\right] &= C\left(\mathbf{s}_i - \mathbf{s}_j\right), \quad \text{for all} \quad \mathbf{s}_i, \mathbf{s}_j \in D \end{aligned}$$

If

$$C\left(\mathbf{s}_i - \mathbf{s}_j\right) = C^o\left(\left\|\mathbf{s}_i - \mathbf{s}_j\right\|\right)$$

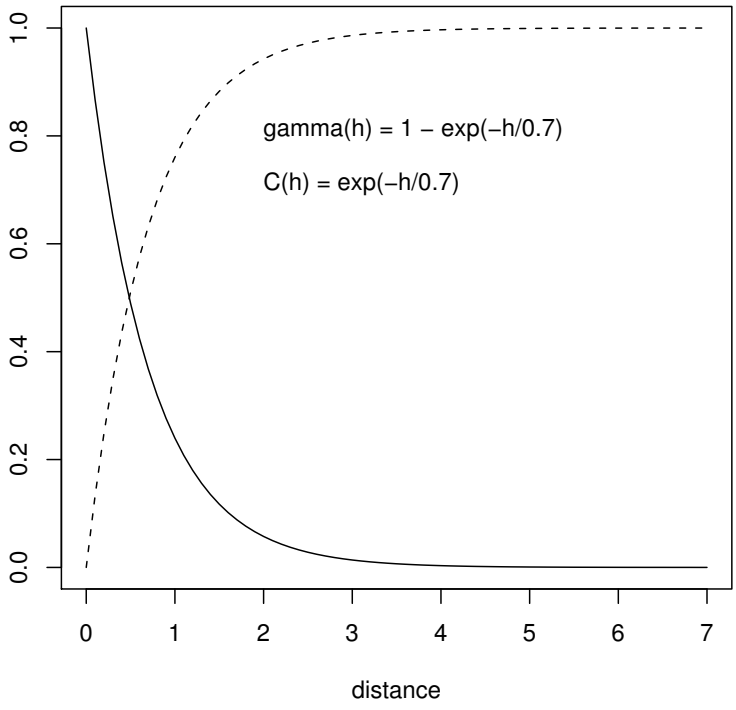
then the process is said to be isotropic, otherwise it is anisotropic. As above, we will often drop the superscript when discussing isotropic covariance functions. Second order stationary processes are intrinsically stationary but the converse is not necessarily true.

- Under second order stationarity we have

$$\gamma(\mathbf{s}_i - \mathbf{s}_j) = C(\mathbf{0}) - C(\mathbf{s}_i - \mathbf{s}_j)$$

where $C(\mathbf{0}) = \mathrm{Var}\left(Z\left(\mathbf{s}\right)\right)$.

A plot of an isotropic exponential covariance function (solid line) and the corresponding isotropic exponential semivariogram (dashed line) is shown below.



- As we saw in Chapter 2 covariance functions of second-order stationary spatial processes satisfy the following properties:
  - $C(\mathbf{0}) \geq 0$
  - $C(\mathbf{h}) = C(-\mathbf{h})$
  - $C(\mathbf{0}) \geq |C(\mathbf{h})|$
  - $C(\mathbf{h}) = \mathrm{Cov}[Z(\mathbf{s}), Z(\mathbf{s}+\mathbf{h})] = \mathrm{Cov}[Z(\mathbf{0}), Z(\mathbf{h})]$
  - If $C_j(\mathbf{h}), j = 1, \cdots, k$ are valid covariance functions then

    $$\sum_{j=1}^{k} b_j C_j(\mathbf{h})$$

    is a valid covariance function if all $b_j \geq 0$ and

    $$\Pi_{j=1}^{k} C_j(\mathbf{h})$$

    is a valid covariance function.
  - If $C(\mathbf{h})$ is a valid covariance function in $\mathcal{R}^d$ it is a valid covariance function in $\mathcal{R}^p$ for $p < d$.

– $C(\mathbf{s}_i - \mathbf{s}_j)$ is a valid covariance function if and only if C is positive definite:

$$\sum_{i=1}^{k}\sum_{j=1}^{k} a_i a_j C(\mathbf{s}_i - \mathbf{s}_j) \geq 0$$

for any finite collection of spatial locations $\mathbf{s}_i, i = 1, \cdots, k$ and real numbers $a_i, i = 1, \cdots, k$.

- The correlation function is defined to be

$$R(\mathbf{s}_i - \mathbf{s}_j) = \frac{C(\mathbf{s}_i - \mathbf{s}_j)}{\sqrt{\text{Var}(Z(\mathbf{s}_i))}\sqrt{\text{Var}(Z(\mathbf{s}_j))}} = \frac{C(\mathbf{s}_i - \mathbf{s}_j)}{C(\mathbf{0})}.$$

R may be referred to as the correlogram. R is called the autocorrelation function in time series analysis.

- *Why the Semivariogram?* You are no doubt familiar with covariances and correlations so why muddy the water with another measure of dependency called the semivariogram? There are several reasons.

1. As we have seen semivariograms exist for a larger class of spatial processes than covariance functions (i.e. the class of intrinsically stationary processes is larger than and contains the class of second order stationary processes).
2. For second order stationary processes

$$\gamma(\mathbf{s}_i - \mathbf{s}_j) \to C(\mathbf{0}).$$

Thus, an estimate of γ at large distances is an estimate of the variance of the process.

3. The most important reason; however, is that the statistical properties of estimators of the semivariogram are better than those of the covariance function. For example, it is easy to see that we need an estimate of the mean in order to estimate the covariance but that is not true for the semivariogram.

- Covariance functions and semivariograms are needed for estimation and prediction. However, they also provide information on the structural properties of spatial processes. Recall the following:

– Sill: If $C(\mathbf{h}) \to 0$ as $\|\mathbf{h}\| \to \infty$ then $\gamma(\mathbf{h}) \to C(\mathbf{0})$. $C(\mathbf{0})$ is referred to as the *sill*.

– Range: The semivariogram may attain the sill or it may approach it asymptotically. The smallest value of $\|\mathbf{h}\|$ at which $\gamma(\mathbf{h}) = C(\mathbf{0})$ is called the *range*. If the semivariogram approaches the sill asymptotically then the range (often referred to as the *practical range* in this case) is generally defined to be

the smallest value of $\|\mathbf{h}\|$ at which the semivariogram is equal to 95% of the sill. The range is interpreted as the distance at which values of Z become uncorrelated, or in the case of the practical range the distance at which spatial correlation becomes unimportant.

Semivariograms of second order stationary processes will always have a sill. However, empirical semivariograms (i.e. $\widehat{\gamma}$) may not exhibit this behavior. This can happen for a number of reasons including:

- The process is intrinsically stationary but not second order stationary.
- The process is not stationary in the mean.
- The (practical) range of the process occurs at a lag longer than that observed in the data.

Another commonly seen feature of empirical semivariograms is a discontinuity at the origin, the *nugget* effect, c_0. In the presence of the nugget effect

$$\text{Var}\left(Z\left(\mathbf{s}\right)\right) = c_0 + \sigma_0^2$$

and σ_0^2 is called the *partial sill*.

The nugget effect can arise from microscale variation (σ_η^2) that is not observable because it occurs at values of $\|\mathbf{h}\|$ smaller than the spacings between the observations. It can also arise as a result of measurement error (σ_ϵ^2). It is common to decompose the nugget effect as

$$c_0 = \sigma_\eta^2 + \sigma_\epsilon^2$$

when discussing it but these two components of variance are oftern not estimable. Microscale variation is hard to deal with empirically as few want to spend time and money collecting enough data at locations so close to one another. Quantifying measurement error requires multiple measurements at the same location and that is not common in many geostatistical applications. Although the determination of how much of the nugget effect is due to microscale variation and how much is due to measurement error is difficult (if not impossible) it is of more than theoretical interest as we will see when we discuss the effect of the sill, range, and nugget effect in spatial prediction later in the course.

- *Parametric Covariance and Semivariogram Models* We will look at some examples of valid models. These are not the only choices but they provide a glimpse of the variety of such models. The versions shown below are isotropic models with $h = \|\mathbf{h}\|$. You may see different parametrizations depending on which source you use. The sill in the models below is

$$C(0) = \lim_{h\to\infty} \gamma(h).$$

1. *Exponential Model*: The semivariogram is

$$\gamma(h) = \begin{cases} 0, & h = 0 \\ c_0 + \sigma_0^2 \left(1 - \exp\left\{-3\frac{h}{\alpha}\right\}\right), & h > 0 \end{cases}$$

where $c_0 \geq 0, \sigma_0^2 \geq 0$, and $\alpha \geq 0$. The nugget effect is c_0, α is the practical range, the partial sill is σ_0^2, and the sill is $C(0) = c_0 + \sigma_0^2$. We can get the covariance function from the relation

$$C(h) = C(0) - \gamma(h)$$

which yields

$$C(h) = \begin{cases} c_0 + \sigma_0^2, & h = 0 \\ \sigma_0^2 \exp\left\{-3\frac{h}{\alpha}\right\}, & h > 0 \end{cases}$$

The exponential model is valid is all dimensions.

2. *Gaussian Model*:

The semivariogram is

$$\gamma(h) = \begin{cases} 0, & h = 0 \\ c_0 + \sigma_0^2 \left(1 - \exp\left\{-3\frac{h^2}{\alpha^2}\right\}\right), & h > 0 \end{cases}$$

where $c_0 \geq 0, \sigma_0^2 \geq 0$, and $\alpha \geq 0$. The nugget effect is c_0, α is the practical range, the partial sill is σ_0^2, and the sill is $C(0) = c_0 + \sigma_0^2$. The covariance function is

$$C(h) = \begin{cases} c_0 + \sigma_0^2, & h = 0 \\ \sigma_0^2 \exp\left\{-3\frac{h^2}{\alpha^2}\right\}, & h > 0 \end{cases}$$

The gaussian model is valid in all dimensions.

3. *Spherical Model*: Typically the spherical semivariogram model is given as

$$\gamma(h) = \begin{cases} 0, & h = 0 \\ c_0 + \sigma_0^2 \left\{(3/2)(h/\alpha) - (1/2)(h/\alpha)^3\right\}, & 0 < h < \alpha \\ c_0 + \sigma_0^2, & h \geq \alpha \end{cases}$$

The corresponding covariance function is

$$C(h) = \begin{cases} c_0 + \sigma_0^2, & h = 0 \\ \sigma_0^2 \left(1 - \left\{(3/2)(h/\alpha) - (1/2)(h/\alpha)^3\right\}\right), & 0 < h < \alpha \\ 0, & h \geq \alpha \end{cases}$$

The spherical model is valid in dimensions 1, 2, and 3.

4. *Rational Quadratic Model*: The semivariogram is

$$\gamma(h) = \begin{cases} 0, & h = 0 \\ c_0 + c_r \left\{ \frac{h^2}{1+h^2/\alpha} \right\}, & h > 0 \end{cases}$$

The corresponding covariance function is

$$C(h) = \begin{cases} c_0 + c_r\alpha, & h = 0 \\ c_r \left(\alpha - \frac{h^2}{1+h^2/\alpha} \right), & h > 0 \end{cases}$$

The parameters are a bit different here. The nugget effect is still c_0, but the sill is $c_0 + c_r\alpha$. The rational quadratic model is valid in all dimensions.

5. *Hole Effect Model*: The semivariogram is

$$\gamma(h) = \begin{cases} 0, & h = 0 \\ c_0 + \sigma_0^2 \left\{ 1 - (\alpha/h) \sin\left(h/\alpha\right) \right\}, & h > 0 \end{cases}$$

The covariance function is

$$C(h) = \begin{cases} c_0 + \sigma_0^2, & h = 0 \\ \sigma_0^2 \left\{ (\alpha/h) \sin\left(h/\alpha\right) \right\}, & h > 0 \end{cases}$$

The sill is $c_0 + \sigma_0^2$ and the practical range is defined to be the lag h "at which the first peak is not greater than 1.05(sill) or the first valley is no less than 0.95(sill).

6. *The Power Model*: The semivariogram is

$$\gamma(h) = \begin{cases} 0, & h = 0 \\ c_0 + \theta h^\lambda, & h > 0 \end{cases}$$

where $c_0 \geq 0, \theta \geq 0$ and $0 \leq \lambda < 2$. Note that as $h \to \infty$ $\gamma(h) \to \infty$, i.e. the sill does not exist. This semivariogram is valid for *intrinsically* stationary processes in all dimensions but it is not valid for second order stationary processes. The *linear model* is a power model with $\lambda = 1$.

The semivariograms discussed above are members of larger families. For example, the exponential and gaussian models are both members of the so-called Matérn Class (see page 143 in Schabenberger and Gotway). Those who work in this area have favorite semivariograms. The only generally consistent opinion seems to be the gaussian family is not a good model for natural spatial processes because it is too smooth (it is the semivariogram of an infinitely differentiable process). My impression is that the spherical model is the favorite among most nonstatistical practitioners.

These parametric families are fit to empirical semivariograms via a variety of methods (more on this below). There has been a lot of work done on nonparametric fitting. This is more difficult than it might at first appear because nonparametric smoothers will not, in general, produce valid semivariograms and covariance functions due to their inability to satisfy the definiteness conditions. I am not aware of any currently available software for nonparametric semivariogram fitting. We are not going to discuss that topic any further here.

- *Estimation of γ and C*: Under intrinsic stationarity we have

$$\gamma(\mathbf{s}_i - \mathbf{s}_j) = \frac{1}{2}\text{Var}\left[Z(\mathbf{s}_i) - Z(\mathbf{s}_j)\right] = \frac{1}{2}E\left[\left(Z(\mathbf{s}_i) - Z(\mathbf{s}_j)\right)^2\right].$$

A natural estimator of γ is Matheron's Estimator:

$$\widehat{\gamma}(\mathbf{h}) = \frac{1}{2|N(\mathbf{h})|}\sum_{N(\mathbf{h})}\left[Z(\mathbf{s}_i) - Z(\mathbf{s}_j)\right]^2$$

where $N(\mathbf{h})$ is the set of all pairs $(\mathbf{s}_i, \mathbf{s}_j)$ with coordinate distance $\mathbf{h} = \mathbf{s}_i - \mathbf{s}_j$ and $|N(\mathbf{h})|$ is the number of such pairs. This works fine for regularly spaced locations but not for irregularly spaced data. Often the coordinate distances are binned and the averaging is carried out over all pairs of points with coordinate distances that fall within a specified bin. Other approaches are also used. The general recommendation is that there should be at least 30 distinct pairs at each lag.

Under intrinsic stationarity $\widehat{\gamma}$ is an unbiased estimator of γ. Also, $\widehat{\gamma}$ satisfies properties that valid semivariograms must possess, with one important exception. For example, $\widehat{\gamma}$ is a symmetric function and $\widehat{\gamma}(\mathbf{0}) = 0$. However, it will not satisfy the conditional negative definiteness condition.

Matheron's estimator is sensitive to extreme observations and extreme observations can impact the estimate at several different lags because each observation is used at many different lags. A robust estimator due to Cressie and Hawkins is

$$\widetilde{\gamma}(\mathbf{h}) = \left(\frac{1}{2}\right)\frac{\left(\frac{1}{|N(\mathbf{h})|}\sum_{N(\mathbf{h})}|Z(\mathbf{s}_i) - Z(\mathbf{s}_j)|^{1/2}\right)^4}{0.457 + 0.494/|N(\mathbf{h})|}.$$

- A natural estimator of the the covariance function is

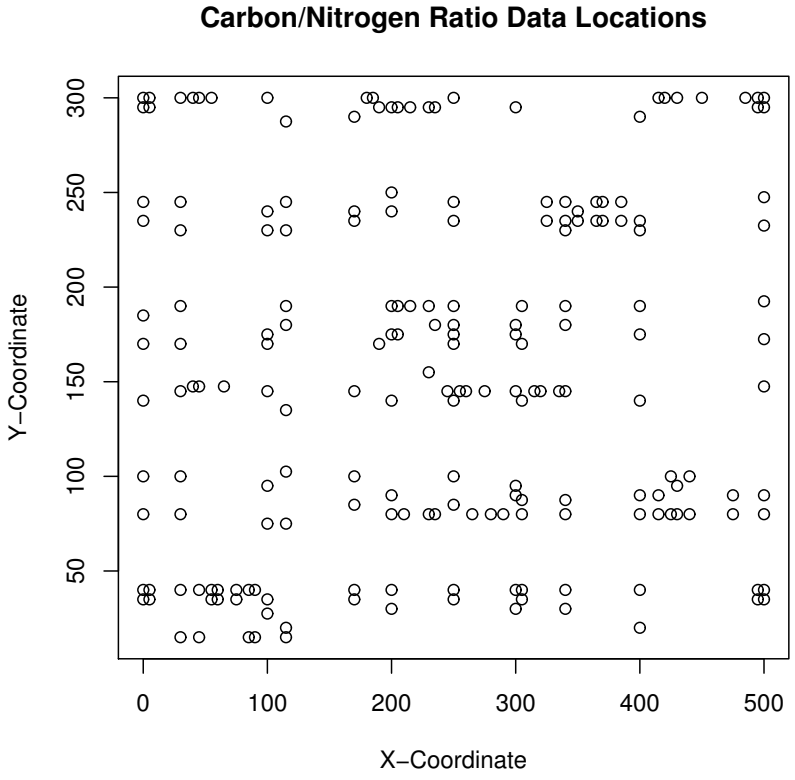
$$\widehat{C}(\mathbf{h}) = \frac{1}{|N(\mathbf{h})|}\sum_{N(\mathbf{h})}\left(Z(\mathbf{s}_i) - \overline{Z}\right)\left(Z(\mathbf{s}_j) - \overline{Z}\right).$$

Estimation of C requires estimation of $E(Z(\mathbf{s}))$ and $\widehat{C}$ will then be a biased estimator of C even under second order stationarity. Other properties of covariance functions are not satisfied (e.g. $\widehat{C}(-\mathbf{h}) \neq \widehat{C}(\mathbf{h})$).

- If a process is not stationary in the mean then both $\widehat{\gamma}$ and $\widehat{C}$ are biased but, in general, $\widehat{\gamma}$ is less effected than $\widehat{C}$.

- There are several functions in R that will estimate semivariograms. These include `vgram` in the library `fields`, `variogram` in Ripley's `spatial` library, `Variogram` in the `nlme` library, `variogram` in the `gstat` library, and `variog` in the `geoR` library. Some of these are easier to use than others. Several of them are designed more for use on residuals from fitted models and are awkward to use on raw data. Some have different scaling defaults, for example, `Variogram` automatically rescales the semivariance values so that the sill, if it exists, is equal to 1. We will look at several of these over the rest of the semester. The following example illustrates the use of `vgram` and `variog`.

  Example: This example is from Schabenberger and Gotway. There are 195 locations in an agricultural field. At each location total carbon and nitrogen percentages were determined. Of interest here is the CN ratio. I loaded the data into R (data frame `CN.dat`*. The spatial locations are shown below.*



  We will look at how to use the various R functions to produce estimate of semivariograms. The text provides us with a few pointers to keep in mind. With irregularly

spaced data it is important to choose the number of bins in such a way that we get 30 or more distinct pairs in each bin. That may require a bit of playing around to determine but in this case 30 to 35 works fine. The authors used 35 and I am going to use about 30. Given the importance of having 30 or more pairs we need to determine the maximum lag we want to work with. One rule of thumb is to choose the maximum to be about half the maximum separation distance observed in the data. We choose 250 feet in this example which is close to what they chose.

The first function we will look at is `vgram`*. The function computes pairwise squared differences as a function of distance and returns either raw values or statistics from binning.*

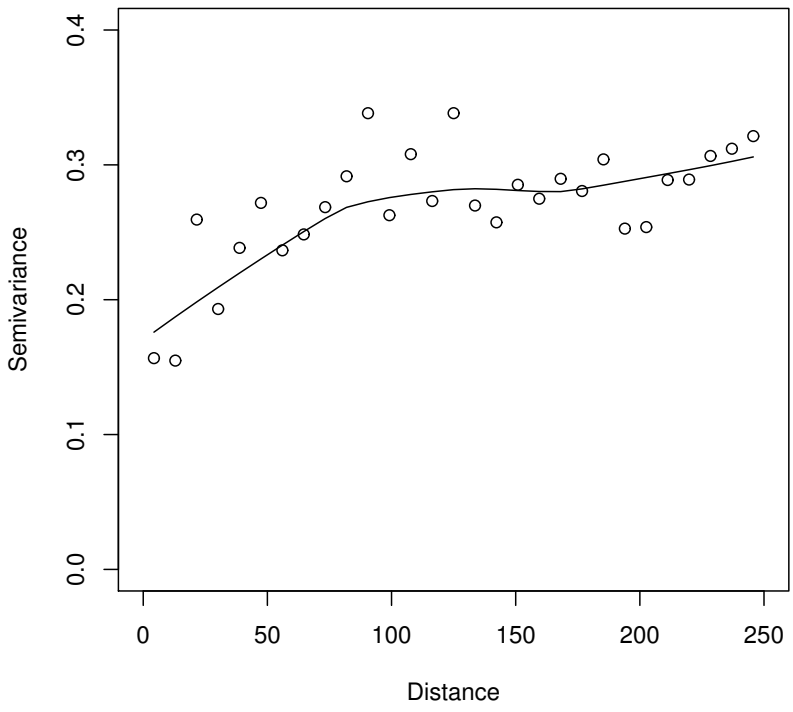
```
dim(CN.dat)
[1] 195   5
> names(CN.dat)
[1] "x"  "y"  "tn" "tc" "cn"

## Put the bins over the first 250 feet of lag distances
brks<-seq(0,250,len=30)
CN.vgram<-vgram(cbind(CN.dat$x,CN.dat$y),CN.dat$cn,N=30,breaks=brks)
```

The first argument is a matrix containing the spatial coordinates and the second is a vector of the observed CN ratios. $N = 30$ *tells* `vgram` *how many bins to use and* `breaks` *tells* `vgram` *where to put the bins (over the first 250 feet of lag distances).*

The results needed for plotting the empirical semivariogram are stored in `CN.vgram$center` *and* `CN.vgram$stats`*. A plot is shown below.*

```
plot(CN.vgram$centers,CN.vgram$stats["mean",],
 xlab="Distance",ylab="Semivariance",main="Matheron's Estimator - CN Data")
```

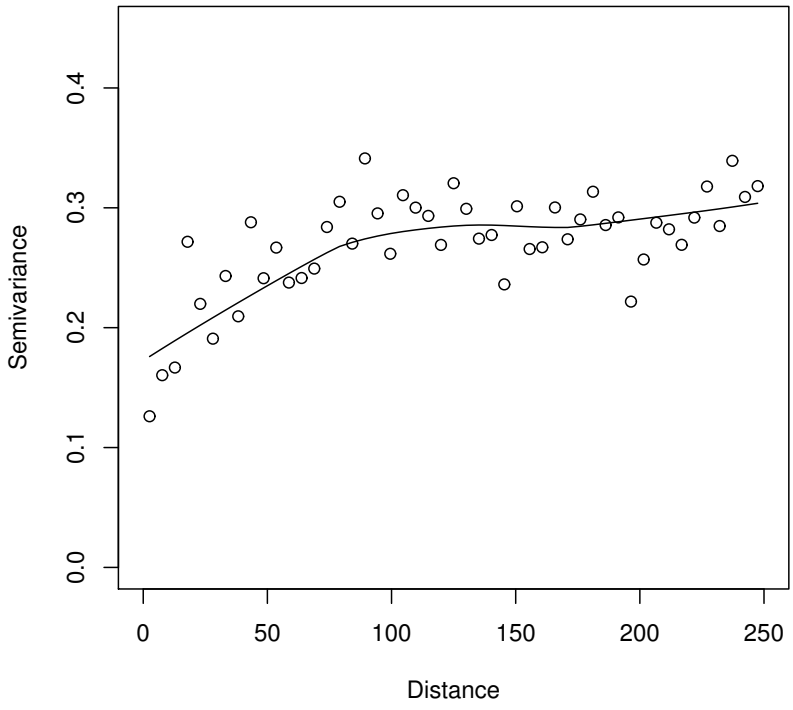
The number of distinct pairs in each lag ranged from 74 to 655. These can be retrieved from `CN.vgram$stats["N",]`*. These all exceed the "magic" cutoff of* 30 *or more distinct pairs at each lag.*



I added the smooth nonparametric curve as an aid to picking out the major structural information only. It is not intended to be used as an actual semivariogram because it almost certainly violates the conditional negative definiteness property. We see a suggestion of a nugget effect of somewhere between 0.10 and 0.15, although the authors do not mention one at this stage (they do model it later). The minimum lag distance is around 4 feet. We see an indication of a sill around 0.25 to 0.30 and a range at around 100 feet. Thus, it appears that the CN ratios are positively correlated up to lag distances of around 100 feet after which they are practically uncorrelated. We can get a bit closer to the origin by specifying more bins.

```
CN.vgram2<-vgram(cbind(CN.dat$x,CN.dat$y),CN.dat$cn,N=50,breaks=seq(0,250,len=50))
plot(c(0,250),c(0,0.45),type="n",xlab="Distance",ylab="Semivariance")
points(CN.vgram2$centers,CN.vgram2$stats["mean",])
lines(lowess(CN.vgram2$centers,CN.vgram2$stats["mean",]))
```

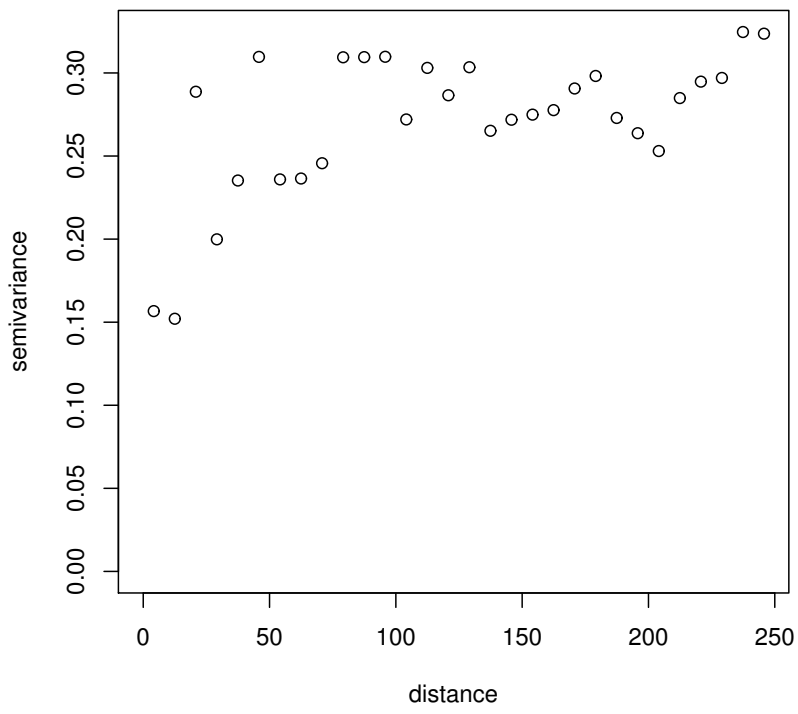
The number of distinct pairs is still pretty good ranging from 52 *to* 306*. The minimum lag distance is now down around* 2.5 *feet. We still see evidence of a nugget effect. Given the nature of the data we might suspect that measurement error is one reason for the nugget effect.*



Although the help page on `vgram` *indicates that it has the ability to calculate a robust estimator it is not clear how to do it. There are no arguments for specifying a robust variogram and no discussion in the details section of the help page. However,* `variog` *does and we will look at that next.*

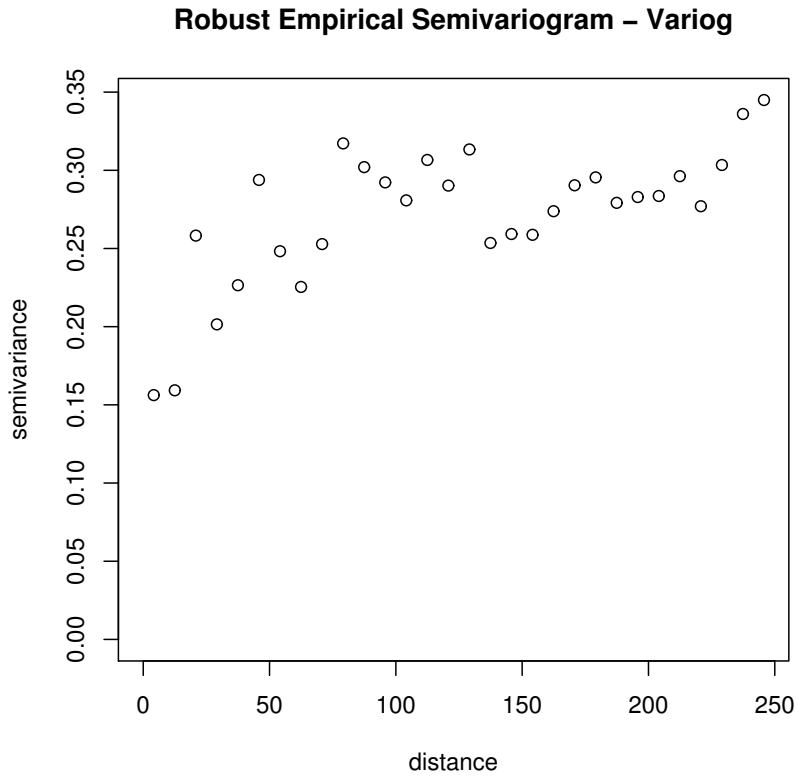
`geoR` *seems to be an interesting package with some interesting functions. The semivariogram function* `variog` *is nicer than* `vgram`*. Additionally,* `geoR` *has the capability of producing simulation envelopes for the empirical semivariogram. The* `variog` *function has a lot of arguments. It is set up nicely to work with data of class* `geodata` *and with data from any data frame. All that is needed are the coordinates and the values of the data recorded at the coordinates. The plot was produced by the following command.*

```
CN.geodata<-as.geodata(CN.dat,coords.col=1:2,data.col=5)
CN.variog<-(variog(CN.geodata,max.dist=250,uvec=30))
plot(CN.variog,main="Empirical Semivariogram - Variog")
```



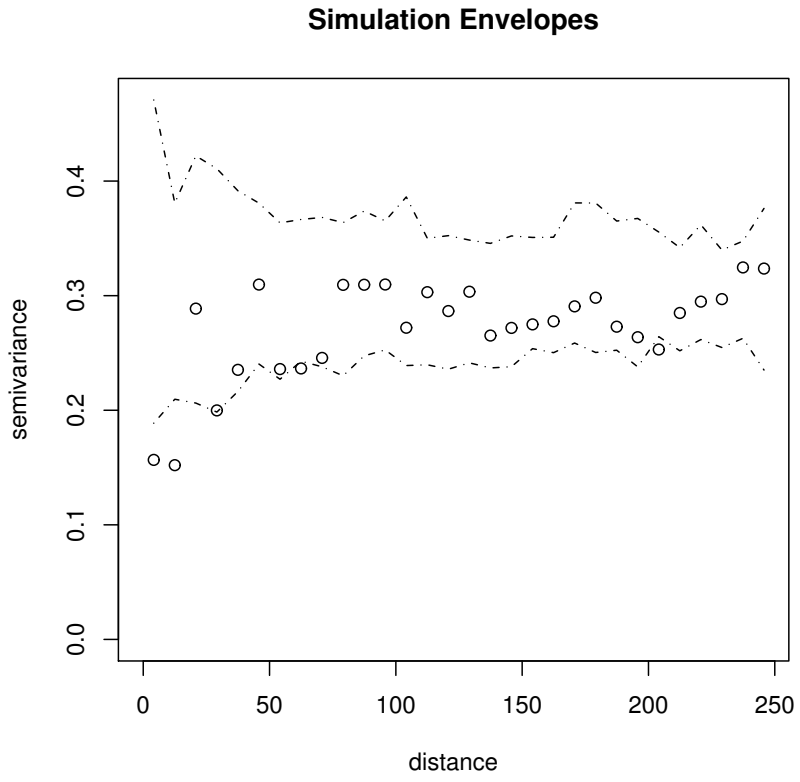
The function also has the ability to estimate the semivariogram using the Cressie and Hawkins robust estimator.

```
CN.variog<-(variog(CN.geodata,max.dist=250,uvec=30,
   estimator.type="modulus"))
plot(CN.variog,main="Robust Empirical Semivariogram - Variog")
```



Simulation envelopes are generated using `variog.mc.env` *as follows.*

```
CN.variog.env<-variog.mc.env(CN.geodata,obj.variog=CN.variog)
plot(CN.variog,envelope=CN.variog.env)
     title(main="Simulation Envelopes")
```



The simulation procedure is described as

```
The envelops (sic) are obtained by permutation.  For each simulations
(sic) data values are randomly allocated to the spatial locations.  The
empirical variogram is computed for each simulation using the same
lags as for the variogram originally computed for the data.  The
envelops (sic) are computed by taking, at each lag, the maximum and
minimum values of the variograms for the simulated data.
```

Simulation envelopes may be more useful when dealing with models.

- *Fitting Semivariograms - Parametric Models*: The most common approach is to fit one of the parametric models described above to an empirical semivariogram, either $\widehat{\gamma}$ or $\widetilde{\gamma}$ (although other empirical estimators may be used as well). One method we will not be using is "fitting by eye" although if you read much in the geostatistical literature you will see that this method is quite common, even today. Some geostatistical software offers this as the only method of fitting. A user sits in front of a computer screen trying various combinations of parameters until the result "looks right". Statisticians generally look upon this method with some skepticism,

although it may be useful when determining good starting values for the numerical fitting routines we discuss below.

1. *Ordinary Least Squares (OLS)*: Recall that the linear regression model with one predictor variable X is

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

and the least squares estimators are the solutions $\left(\widehat{\beta}_0, \widehat{\beta}_1\right)$ which minimize

$$Q = \sum_{i=1}^{n} \left[Y_i - (\beta_0 + \beta_1 x_i)\right]^2$$

or in matrix notation

$$Q = (\mathbf{Y} - \boldsymbol{X}\boldsymbol{\beta})' (\mathbf{Y} - \boldsymbol{X}\boldsymbol{\beta})$$

with solutions

$$\widehat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{X}'\mathbf{Y}.$$

In nonlinear regression we have

$$Y_i = f(\mathbf{X}_i, \boldsymbol{\theta}) + \epsilon_i$$

where $\boldsymbol{\theta}$ is a vector of parameters and

$$\mathbf{X}_i = \begin{bmatrix} X_{i1} \\ X_{i2} \\ \vdots \\ X_{ik} \end{bmatrix}.$$

We seek solutions $\widehat{\boldsymbol{\theta}}$ which minimize

$$Q = \sum_{i=1}^{n} \left[Y_i - f(\mathbf{X}_i, \boldsymbol{\theta})\right]^2.$$

Generally a closed form solution does not exist and a numerical procedure (e.g. the Gauss-Newton method) is used. For the Gauss-Newton method we "linearize" the problem by approximating $f(\mathbf{X}_i, \boldsymbol{\theta})$ by its first order Taylor Series Expansion:

$$f(\mathbf{X}_i, \boldsymbol{\theta}) \approx f\left(\mathbf{X}_i, \boldsymbol{\theta}^{(0)}\right) + \sum_{j=1}^{p} \left[\frac{\partial f(\mathbf{X}_i, \boldsymbol{\theta})}{\partial \theta_j}\right]_{\boldsymbol{\theta} = \boldsymbol{\theta}^{(0)}} \left(\theta_j - \theta_j^{(0)}\right)$$

where $\boldsymbol{\theta}^{(0)}$ is a vector of parameter starting values. The original nonlinear model

$$Y_i = f\left(\mathbf{X}_i, \boldsymbol{\theta}\right) + \epsilon_i$$

is thus approximated by

$$Y_i = f\left(\mathbf{X}_i, \boldsymbol{\theta}^{(0)}\right) + \sum_{j=1}^{p} \left[\frac{\partial f\left(\mathbf{X}_i, \boldsymbol{\theta}\right)}{\partial \theta_j}\right]_{\boldsymbol{\theta} = \boldsymbol{\theta}^{(0)}} \left(\theta_j - \theta_j^{(0)}\right) + \epsilon_i$$

which is a linear model. We actually work with the no-intercept linear model

$$Y_i - f\left(\mathbf{X}_i, \boldsymbol{\theta}^{(0)}\right) = \sum_{j=1}^{p} \left[\frac{\partial f\left(\mathbf{X}_i, \boldsymbol{\theta}\right)}{\partial \theta_j}\right]_{\boldsymbol{\theta} = \boldsymbol{\theta}^{(0)}} \left(\theta_j - \theta_j^{(0)}\right) + \epsilon_i$$

The solution is found iteratively, i.e. the first step results in a "solution" $\boldsymbol{\theta}^{(1)}$ which serves as new "starting values" for the next iteration. The sum of squared errors at each iteration

$$SSE^{(m)} = \sum_{i=1}^{n} \left[Y_i - f\left(\mathbf{X}_i, \boldsymbol{\theta}^{(m)}\right)\right]^2$$

is computed and iteration ceases when that term stabilizes.

In the context of fitting semivariograms our "data" are either the individual terms

$$\left(Z\left(\mathbf{s}_i\right) - Z\left(\mathbf{s}_j\right)\right)^2$$

or the empirical semivariogram values $\widehat{\gamma}(\mathbf{h})$ or some other version. There has been some discussion about which is better to use but most practitioners (and most software packages) use Matheron's estimator. In this latter case the data vector is then

$$\widehat{\boldsymbol{\gamma}}(\mathbf{h}) = \left[\widehat{\gamma}(\mathbf{h}_1), \widehat{\gamma}(\mathbf{h}_2), \cdots, \widehat{\gamma}(\mathbf{h}_k)\right]'$$

and the nonlinear function being fit is a parametric semivariogram (e.g. exponential, rational quadratic, etc.) $\gamma\left(\mathbf{h}, \boldsymbol{\theta}\right)$ where $\boldsymbol{\theta}$ is the vector containing the nugget, sill (or partial sill), and (practical) range parameters (or some subset of those).

2. *Weighted Least Squares (WLS)*: Ordinary least squares assumes that $\text{Var}\left(\widehat{\gamma}(\mathbf{h})\right)$ is constant for all $\mathbf{h}$. Weighted least squares allows us to lighten up on that assumption. We specify a variance-covariance matrix for the $\widehat{\gamma}$'s that has diagonal elements $\text{Var}\left(\widehat{\gamma}(\mathbf{h})\right)$ different for different $\mathbf{h}$. An approximation to this variance is

$$\text{Var}\left(\widehat{\gamma}(\mathbf{h})\right) \approx 2\left[\frac{\gamma\left(\mathbf{h}_j, \boldsymbol{\theta}\right)^2}{\left|N\left(\mathbf{h}_j\right)\right|}\right]$$

which obviously has to be estimated. The solution we are looking for is the $\widehat{\boldsymbol{\theta}}$ which minimizes

$$\begin{aligned} Q &= (\widehat{\boldsymbol{\gamma}}(\mathbf{h}) - \boldsymbol{\gamma}(\mathbf{h}, \boldsymbol{\theta}\ )^{'}\, \mathbf{W}\,(\boldsymbol{\theta})^{-1}\,(\widehat{\boldsymbol{\gamma}}(\mathbf{h}) - \boldsymbol{\gamma}(\mathbf{h}, \boldsymbol{\theta}\ ) \\ &= \sum_{j=1}^{k} \frac{|N\,(\mathbf{h}_j)\,|}{2\gamma\,(\mathbf{h}_j, \boldsymbol{\theta})^2}\,\{\widehat{\gamma}\,(\mathbf{h}_j) - \gamma\,(\mathbf{h}_m, \boldsymbol{\theta})\}^2 \end{aligned}$$

3. *Generalized Least Squares*: The weight matrix now has off diagonal elements, i.e. it has terms for the

$$\mathrm{Cov}\,(\widehat{\gamma}\,(\mathbf{h}_i)\,, \widehat{\gamma}\,(\mathbf{h}_j))\,.$$

   The generalized weight matrix is denoted $\mathbf{R}\,(\boldsymbol{\theta})$ and the goal is now to find $\widehat{\boldsymbol{\theta}}$ which minimizes

$$Q = (\widehat{\boldsymbol{\gamma}}(\mathbf{h}) - \boldsymbol{\gamma}(\mathbf{h}, \boldsymbol{\theta}\ )^{'}\, \mathbf{R}\,(\boldsymbol{\theta})^{-1}\,(\widehat{\boldsymbol{\gamma}}(\mathbf{h}) - \boldsymbol{\gamma}(\mathbf{h}, \boldsymbol{\theta}\ )\,.$$

4. *Maximum Likelihood Estimation (MLE)*: If we are willing to assume that $Z\,(\mathbf{s}) \sim G\,(\mu\mathbf{1}, \boldsymbol{\Sigma}\,(\boldsymbol{\theta}))$ then we can estimate $\boldsymbol{\theta}$ by maximum likelihood estimation. The negative of twice the log likelihood is

$$\begin{aligned} \phi\,(\boldsymbol{\mu}, \boldsymbol{\theta}, \mathbf{Z}(\mathbf{s})) &= \ln\{|\boldsymbol{\Sigma}\,(\boldsymbol{\theta})\,|\} + n\ln\,(2\pi) \\ &\quad + (\mathbf{Z}\,(\mathbf{s}) - \mathbf{1}\mu)^{'}\,\boldsymbol{\Sigma}\,(\boldsymbol{\theta})^{-1}\,(\mathbf{Z}\,(\mathbf{s}) - \mathbf{1}\mu) \end{aligned}$$

   We have to play games with μ. Assuming $\boldsymbol{\theta}$ is know the maximum likelihood estimator of μ is

$$\widetilde{\mu} = \left(\mathbf{1}^{'}\boldsymbol{\Sigma}\,(\boldsymbol{\theta})^{-1}\,\mathbf{1}\right)^{-1}\mathbf{1}^{'}\boldsymbol{\Sigma}\,(\boldsymbol{\theta})^{-1}\,\mathbf{Z}(\mathbf{s})$$

   and this expression is plugged back into the minus 2 log likelihood ϕ (this is an example of a profile likelihood). After substitutiion, ϕ is a function only of $\boldsymbol{\theta}$ from which maximum likelihood estimators $\widehat{\boldsymbol{\theta}}_{ml}$ are determined. As you know maximum likelihood estimators have nice statistical properties (although they tend to be biased).

5. *Restricted Maximum Likelihood Estimation (REML)*: The mean μ causes some problems in maximum likelihood estimation. REML is a method of dealing with that by taking out the mean and maximizing the resulting restricted likelihood equation. We will skip the details noting only that the goal is find a matrix $\mathbf{K}$ such that

$$E\,[\mathbf{K}Z(\mathbf{s})] = \mathbf{0}$$

   for all $\mathbf{s}$. The matrix $\mathbf{K}$ is incorporated into the likelihood equations.

There are other approaches including generalized estimating equations (GEE) and composite likelihoods for estimating $\boldsymbol{\theta}$ but we will skip over those. The `geoR` package

has the ability to fit semivariogram (covariance) models by OLS, WLS, MLE, and REML and those are the 4 methods we will work with in this course.

Example: We will use the carbon-nitrogen ratio data. We will fit the exponential model using the `variofit` *function in* **R**. *The code is shown below. It is provided here to give you an idea of what to expect. The OLS and WLS results are quick but the MLE and REML results can take a while. It is a good idea to play around with the starting values a little also to see how stable the results are. The* phi *parameter is the range parameter and must be multiplied by* 3 *to get the practical range.*

```
# OLS with no nugget
CN.ols<-variofit(CN.variog,ini=c(0.3,15),fix.nugget=T,nugget=0,
       cov.model="exponential",weights="equal")
CN.ols
variofit: model parameters estimated by OLS (ordinary least squares):
covariance model is: exponential
fixed value for tausq =  0
parameter estimates:
sigmasq     phi
  0.285  13.045
Practical Range with cor=0.05 for asymptotic range: 39.07935

# WLS with no nugget
CN.wls<-variofit(CN.variog,ini=c(0.3,15),fix.nugget=T,nugget=0,
    cov.model="exponential")
CN.wls
variofit: model parameters estimated by WLS (weighted least squares):
covariance model is: exponential
fixed value for tausq =  0
parameter estimates:
sigmasq     phi
 0.2883 19.1673
Practical Range with cor=0.05 for asymptotic range: 57.42012

# OLS and WLS with nugget
CN.ols.f<-variofit(CN.variog,ini=c(0.17,15),fix.nugget=F,
 nugget=0.13,cov.model="exponential",weights="equal")
CN.ols.f
variofit: model parameters estimated by OLS (ordinary least squares):
covariance model is: exponential
parameter estimates:
  tausq sigmasq     phi
 0.1369  0.1566 34.4202
```

```
Practical Range with cor=0.05 for asymptotic range: 103.1137

variofit: minimised sum of squares = 0.0202


CN.wls.f<-variofit(CN.variog,ini=c(0.17,15),fix.nugget=F,nugget=0.13,
  cov.model="exponential")
CN.wls.f
variofit: model parameters estimated by WLS (weighted least squares):
covariance model is: exponential
parameter estimates:
  tausq sigmasq     phi
 0.1333  0.1595 36.9044
Practical Range with cor=0.05 for asymptotic range: 110.5556

variofit: minimised weighted sum of squares = 6.7529

# MLE and REML with no nugget

CN.mle<-likfit(CN.geodata,ini=c(0.17,15),fix.nugget=T,nugget=0,
 cov.model="exponential")
CN.mle
likfit: estimated model parameters:
     beta   sigmasq       phi
"10.8175" " 0.3137" "13.4729"
Practical Range with cor=0.05 for asymptotic range: 40.36132

likfit: maximised log-likelihood = -137.1
CN.reml<-likfit(CN.geodata,ini=c(0.30,15),fix.nugget=T,nugget=0,
 cov.model="exponential",lik.method="RML")
CN.reml
likfit: estimated model parameters:
     beta   sigmasq       phi
"10.8183" " 0.3187" "13.8664"
Practical Range with cor=0.05 for asymptotic range: 41.54004

likfit: maximised log-likelihood = -136.3


MLE and REML with nugget

CN.reml.f<-likfit(CN.geodata,ini=c(0.17,15),fix.nugget=F,nugget=0.13,
```

```
 cov.model="exponential",lik.method="RML")
CN.reml.f
likfit: estimated model parameters:
    beta     tausq    sigmasq        phi
"10.8594" " 0.1180" " 0.2159" "57.0517"
Practical Range with cor=0.05 for asymptotic range: 170.9117

likfit: maximised log-likelihood = -129.8

CN.mle.f<-likfit(CN.geodata,ini=c(0.17,15),fix.nugget=F,nugget=0.13,
  cov.model="exponential")
CN.mle.f
likfit: estimated model parameters:
    beta      tausq    sigmasq        phi
"10.8507"  " 0.1132"  " 0.2023"  "47.0144"
Practical Range with cor=0.05 for asymptotic range: 140.8426

likfit: maximised log-likelihood = -131.2
```

We can summarize the results as follows. Note the effect of the including a nugget effect on the practical range.

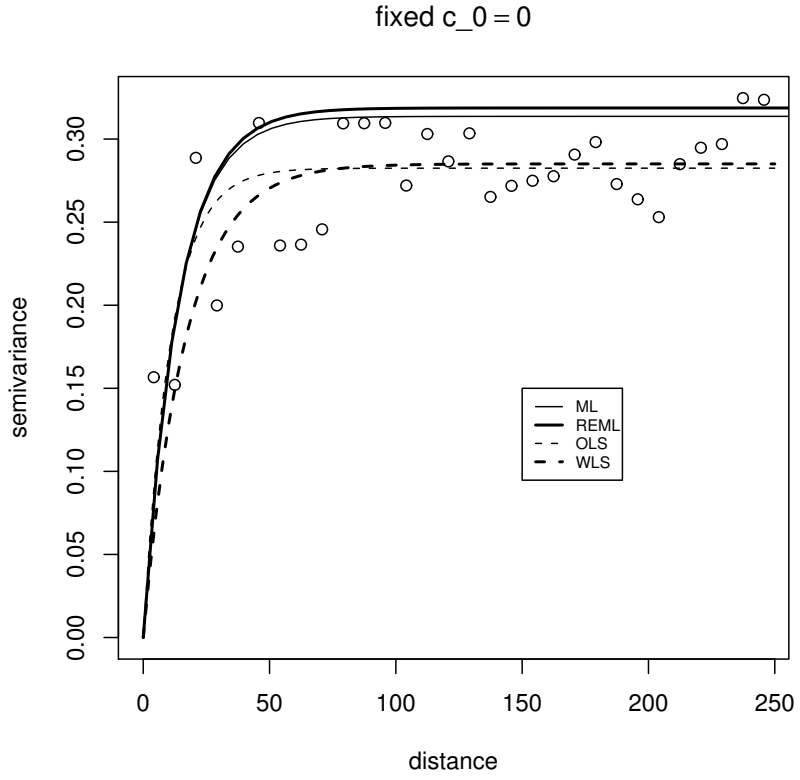
Method	*Nugget*	*(Partial) Sill*	*Practical Range*
OLS	*0*	*0.285*	*39.079*
OLS	*0.1369*	*0.1566*	*103.114*
WLS	*0*	*0.2883*	*57.420*
WLS	*0.1333*	*0.1595*	*110.5556*
MLE	*0*	*0.3137*	*40.3613*
MLE	*0.1132*	*0.2023*	*140.843*
REML	*0*	*0.3187*	*41.540*
REML	*0.1180*	*0.2159*	*170.912*

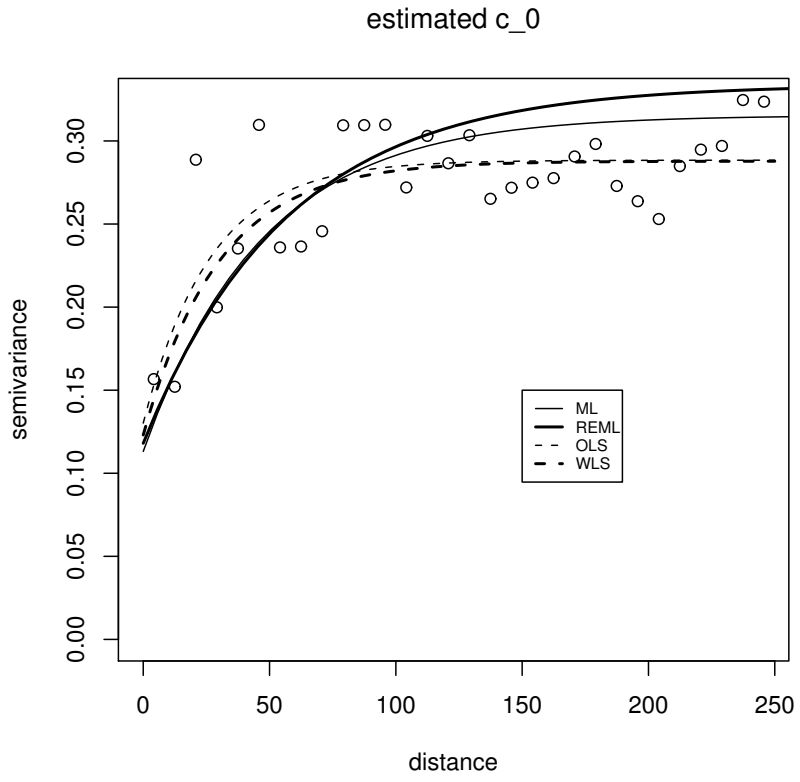
We can test for the importance of a nugget effect using the likelihood based models. The `likfit` *function returns the maximized log likelihood. The loglikelihood for the model with no nugget effect is* -137.1 *and for the model with a nugget effect it is* -131.2*. The effect of adding the parameter is*

$$-2\left(-137.1+131.2\right)=11.8.$$

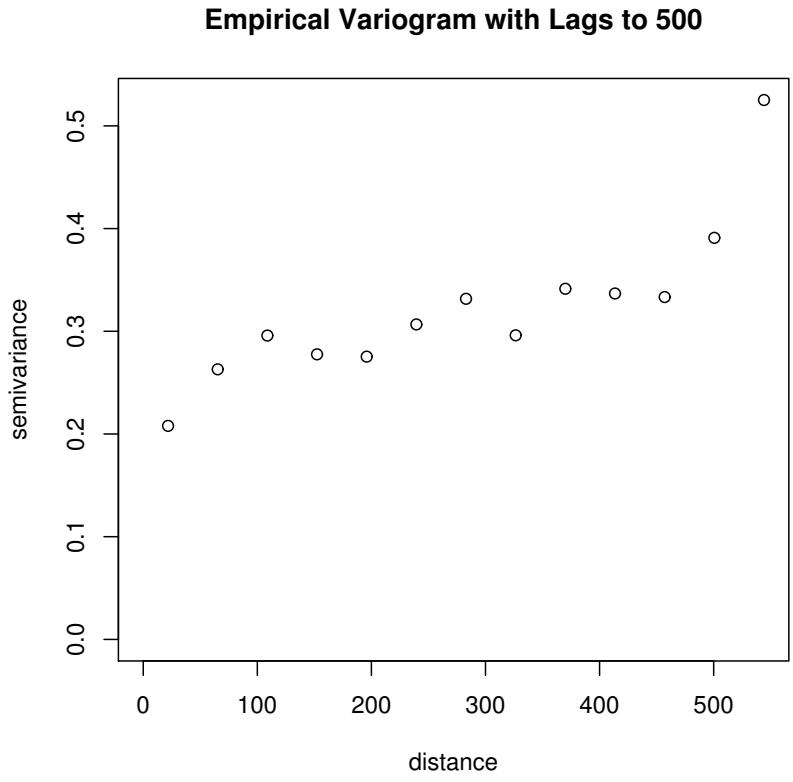
Under a null hypothesis that the nugget effect is 0 the test statistic has a chi-squared distribution with 1 df which yields a p value of 0.0006*, strong evidence for the presence of a nugget effect. We could carry out the same test using the results from the REML analysis and would get pretty much the same answer. We could also do AIC based comparisons. The* ΔAIC *in this case would be 9.8. If you do the AIC based approach remember to include the* `beta` *parameter (which you can ignore for the estimates of nugget, sill, and range). Note that one could also do AIC based comparisons of different models; e.g. spherical or exponential covariance (semivariogram) function. The test is strongly dependent on the assumptions and the MLE and REML results are technically valid only under an assumption of normality (a Gaussian random field) which should at least be assessed. The CN ratio data actually looks pretty good.*

Plots of the results are shown below.





Visual assessments almost always suggest that OLS/WLS do better than $MLE/REML$. This is guaranteed by the fitting criteria. Also, the $MLE/REML$ results are impacted by the effects at lags longer than those seen in the plot because those methods use all the data. There can be a lot of instability at those longer lags (see the plot below).



Schabenberger and Gotway obviously like the $MLE/REML$ methods the best because theoretically they allow for some inference, e.g. they provide valid standard errors if the assumptions are met. As they point out the standard errors from the least squares estimates will not be valid and it is not possible to test for the the presence of a nugget effect.

- Anisotropy A geostatistical spatial process is said to be *anisotropic* when the spatial variability varies with direction. There are 2 types: geometric anisotropy and zonal anisotropy.

    1. Geometric Anisotropy - Mathematically a geometrically anisotropic process is one in which the anisotropy can be transformed to isotropy by a simple linear transformation of the coordinates. For a second order stationary process geometric anisotropy would be characterized by directional semivariograms with different ranges but the same sills.

       Suppose we have a second order stationary spatial process with elliptical isocorrelation contours. Let (x_1, y_1) be the initial rectangular coordinates, and let θ be the angle between the x_1 axis and the major axis of the ellipse. The ratio of the range on the major axis to the range on the minor axis of the ellipse is called the *anisotropy ratio* or sometimes the *ratio of affinity*. Correction for geometrical anisotropy has 2 (optionally 3 steps):

       (a) First we rotate the original coordinates through the angle θ so that the original axes parallel the axes of the ellipse. This yields a new coordinate system with coordinates:

       $$\begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \end{bmatrix}$$

       (b) Now we transform the ellipse into a circle whose radius is equal to the range parameter associated with the major axis of the ellipse:

       $$\begin{bmatrix} x_3 \\ y_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & \lambda \end{bmatrix} \begin{bmatrix} x_2 \\ y_2 \end{bmatrix}$$

       (c) If desired (and this step is really not required) we can rotate these coordinates through an angle of $-\theta$ to get the original orientation.

       Of course this can all be accomplished in a single step:

       $$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos(-\theta) & -\sin(-\theta) \\ \sin(-\theta) & \cos(-\theta) \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & \lambda \end{bmatrix} \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \end{bmatrix}$$

       So one approach to correcting geometrical anisotropy would be to fit semivariograms through several different directions to identify the ellipse (especially the major and minor axes of the ellipse). Estimate a range parameter for the 2 axes. Estimate the angle θ. Generate a new coordinate system using the transformation described above. `geoR` has a function (`coords.aniso`) that will do this. An isotropic semivariogram could then be fit to the new coordinates.

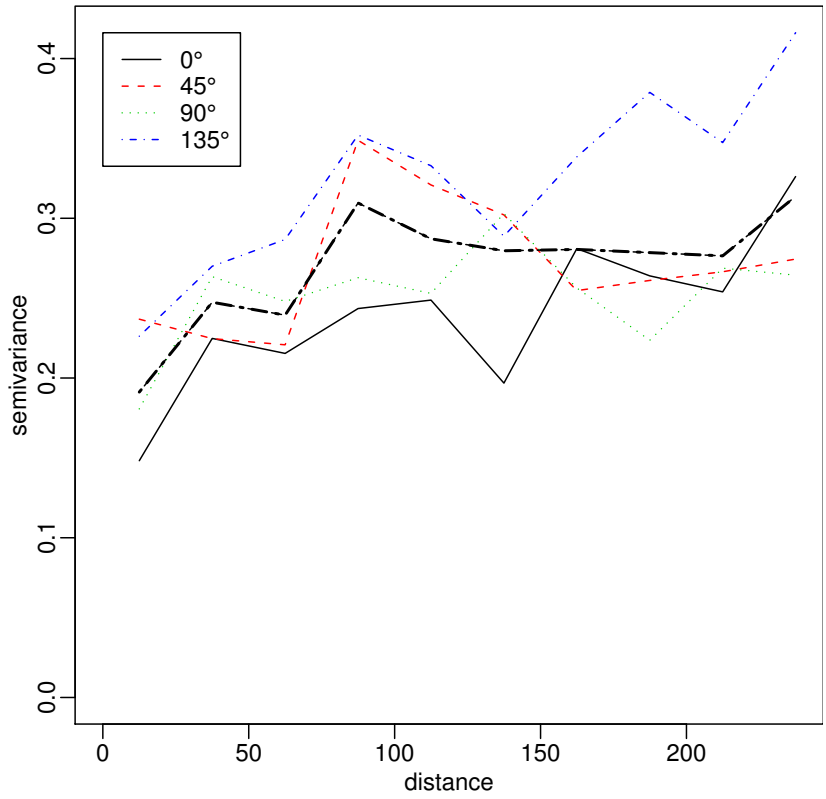
    2. Zonal Anisotropy - Basically if an anisotropic process is not geometrically anisotropic then it is zonally anisotropic. Typically you will see not only different ranges but different sills. The reduction to isotropy is accomplished by

nesting, or by expressing the final semivariogram as a sum of other semivariograms which describe the behavior in different directions. Schabenberger and Gotway give a generic example of a semivariogram that is composed of an isotropic semivariogram and a second semivariogram that depends only on the lag distance in the direction θ of the larger sill. This model may be appropriate when the shorter range is associated with the smaller sill. Zonal anisotropy is somewhat harder to deal with than geometric anisotropy.

As a general rule anisotropy should not be a surprise. The subject matter expert will usually have some idea of whether or not anisotropy will be present. For example, if one is investigating wind born deposits of a pollutant down wind of a smoke stack then geometric anisotropy is to be expected.

Example: `geoR` *has the capability of fitting empirical semivariograms through different directions. A good starting point is to examine the spatial correlation structure in 4 directions* $0°, 45°, 90°, 135°$*. One problem with choosing different directions is that the "sample size" is reduced so directional semivariograms will tend to be noisier than isotropic semivariograms. We examine the CN data to search for evidence of anisotropy.*

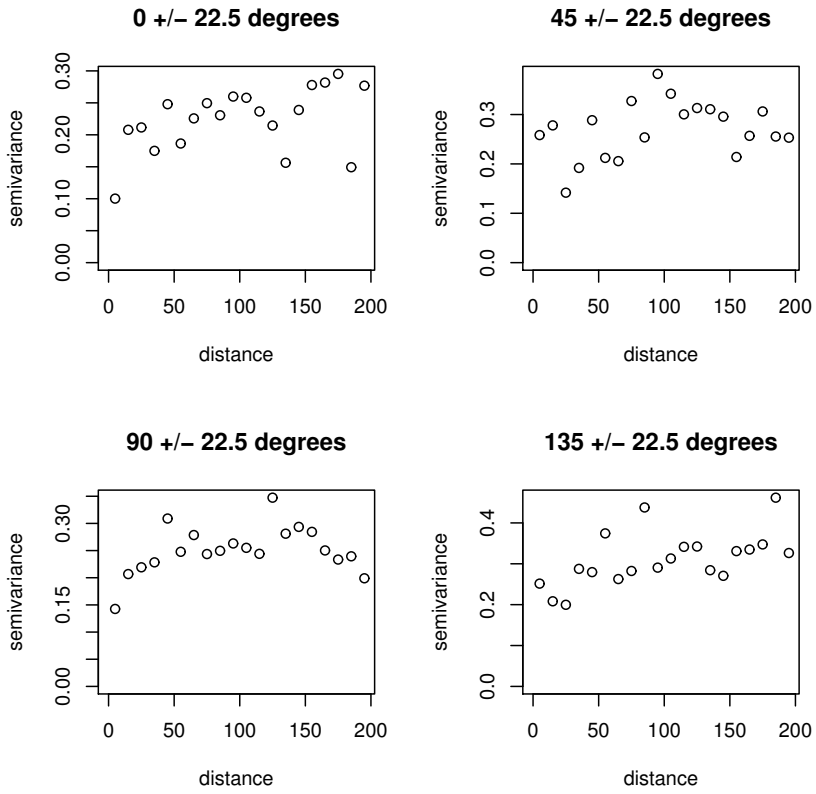
```
CN.aniso<-variog4(CN.geodata,max.dist=200,uvec=10)
plot(CN.aniso)
```



I originally tried to have a few more bins but the resulting plot was very erratic. The 4 angles are the default directions. In addition, the angle tolerance default is $\pm 22.5°$ *which ensures nonoverlapping segments. The bolder line is the omnidirectional semivariogram. Schabenberger and Gotway concluded that there is little evidence of anisotropy here.*

Below is another plot looking at directional semivariograms for these data. I increased the number of bins for these. The results do seem a bit more equivocal. So, does the automatic "smoothing" `geoR` *does in the first plot above shed light on the structure or is it more likely to highlight structure that is not really there.*

```
CN.aniso<-variog4(CN.geodata,uvec=20,max.dist=200)
par(mfrow=c(2,2))
plot(CN.aniso$"0",main="0 +/- 22.5 degrees")
plot(CN.aniso$"45",main="45 +/- 22.5 degrees")
plot(CN.aniso$"90",main="90 +/- 22.5 degrees")
plot(CN.aniso$"135",main="135 +/- 22.5 degrees")
```



In practice I would, as a statistician, be sure to discuss possible anisotropy with the subject matter expert. They really should have some a priori idea of the type of anisotropy likely to be present and, in many cases, the directions in which the anisotropy should manifest itself. We will have more to say about this anisotropy later.

- In this section we have learned some basics about modeling semivariograms (and covariograms and correlograms). We will now turn to using these parametric models in spatial prediction (kriging) and spatial regression models.