University of California, Los Angeles
Department of Statistics

Statistics C173/C273 **Instructor: Nicolas Christou**

More on the variogram

Variogram model parameters:
The various parameters of the variogram model are:

1. Nugget Efffect (c_0):
   If we stand by the assumption that sample values are measured precisely and accurately then the semi-variogram model must have a value of zero at zero distance. It is like calculating the difference of $Z(s)$ with itself. That is,

   $$\gamma(0) = 0$$

   The term nugget effect (or nugget variance) was introduced on the basis of the interpretation of gold mineralization. It was suggested by Matheron (1962) and it is believed that microscale variation (small nuggets) is causing a discontinuity at the origin.

2. Range (α):
   As the separation distance increases the value of the variogram increases as well. However, after a certain distance the variogram reaches a plateau. The distance at which the variogram reaches a plateau is the range.

   We generally interpret the range of influence as that distance beyond which pairs of sample values are unrrelated. Beyond the range the variogram remains essentially constant.

3. Sill ($c_0 + c_1$):
   It is the variogram value for separation distances $h \geq \alpha$.

Outline of spatial continuity analysis:

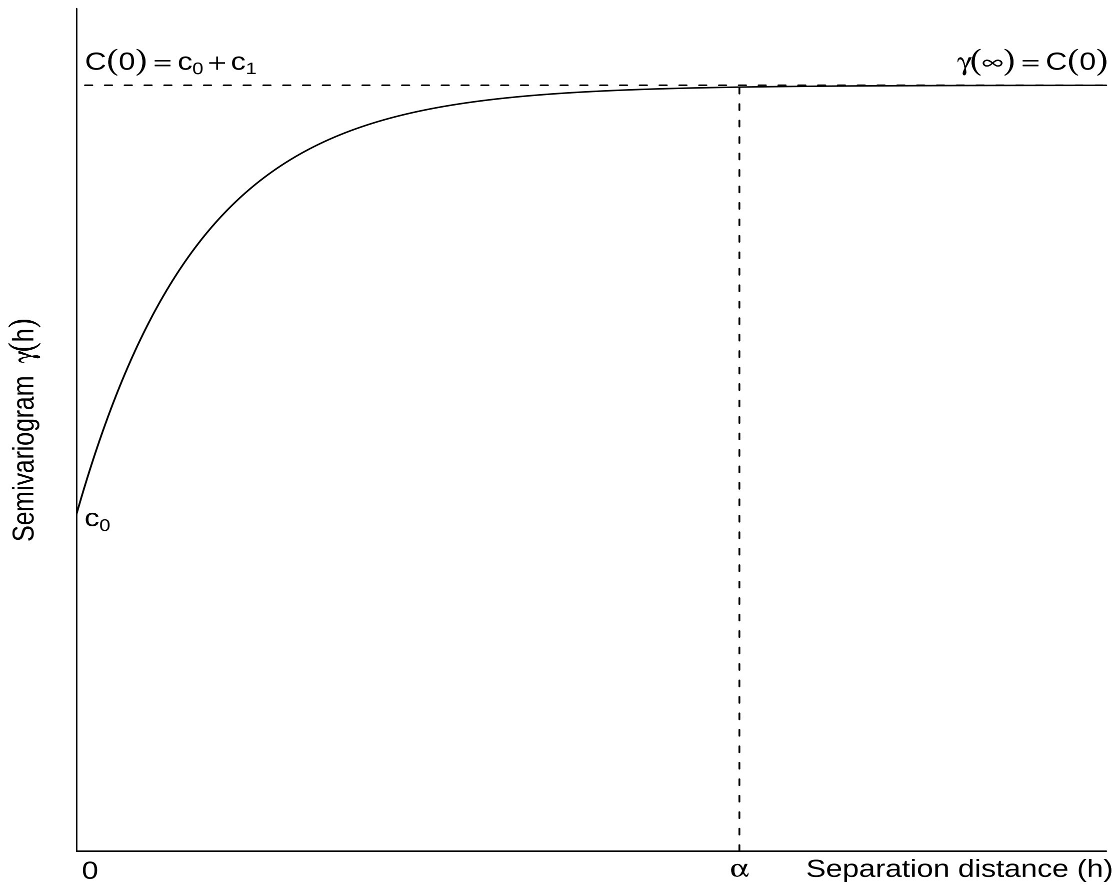
1. We begin usually with the calculation of an *omnidirectional* variogram. With all possible directions combine in a single variogram only the separation distance is important. The omnidirectional variogram can be thought as the average variogram of all directions. Strictly speaking is not the average because in one direction we may have more pairs than other directions.

2. The second step is to explore anisotropy by calculating the *directional* variograms for different directions (one at a time). In many spatial data the direction of anisotropy may be determined by the nature of the problem. For example, if we analyze airborne pollutants, the wind direction may be an important factor in the calculation of the variogram.

3. Once we decide which directional variogram we want to calculate, we must choose the distance parameters. There are two distance parameters. The first one is the separation distance (h). If the samples form approximately a grid, then the grid distance can be a good choice for h. If the samples do not form a regular grid the separation distance is chosen to be equal to the average distance of neighboring samples. The second distance parameter is the tolerance we allow on h. This will give us enough pairs for the variogram calculation. The common choice for the tolerance is half of the separation distance (h). For example, if we use $h = 10\ m$ then we will use all the points that fall between 5 m and 15 m, etc.

4. Another parameter that we need to choose is the *angular* tolerance. When we calculate directional variograms ideally we want to use small angular tolerances so that the direction is preserved. However, many times we need to choose an angular tolerance large enough to produce enough number of pairs for the variogram calculation. We can try different angular tolerances and use the smallest one that produces good results.

Exponential variogram with nugget:



Example:
Suppose that Z is a second order stationary process with $E[Z(s)] = 0$ and with spherical semivariogram:

$$\gamma(h;\boldsymbol{\theta}) = \begin{cases} 0, & h = 0 \\ 0.5 + 4\left(\frac{3}{2}\left(\frac{h}{30}\right) - \frac{1}{2}\left(\frac{h}{30}\right)^3\right), & 0 < h \leq \alpha \\ 4.5, & h > \alpha \end{cases}$$

a. What is the sill of Z?

b. What is the nugget effect of Z?

c. Draw this variogram (approximately). Make sure you place some important number on the graph.

d. Compute $\gamma(5)$.

e. Write the covariance function $C(h;\boldsymbol{\theta})$ that corresponds to the spherical semivariogram above.

More examples:
We will use again the data from the previous tutorial:

44		40	42	40	39	37	36	
42		43	42	39	39	41	40	38
37	37	37	35	38	37	37	33	34
35	38		35	37	36	36	35	
36	35	36	35	34	33	32	29	28
38	37	35		30		29	30	32

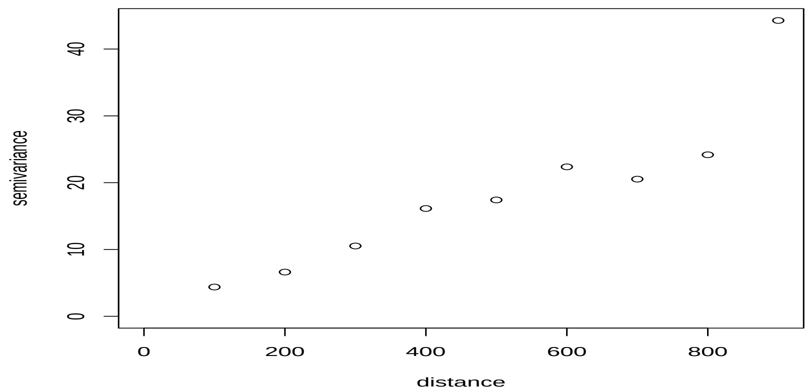
The distance between two points is $100m$ (north-south or east-west).
We read the file and convert the data into geodata:

```
> a <- read.table("http://www.stat.ucla.edu/~nchristo/data_var.txt",
  header=TRUE)
> b <- as.geodata(a)
```

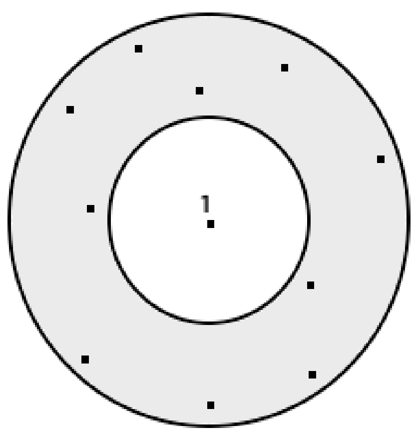
Understand the omnidirectional variogram:
We compute the omnidirectional variogram as follows:

```
> var1 <- variog(b, max.dist=900, uvec=seq(0, 900, by=100))
> var1$u
[1] 100 200 300 400 500 600 700 800 900
> var1$v
[1]  4.381481  6.619760 10.525568 16.134000 17.414179 22.369369
[7] 20.531250 24.176471 44.277778
> var1$n
[1] 135 167 176 250 134 111  64  34   9
```

Here is the plot:



How was the omnidirectional variogram computed?
The lag distances are $100\ m, 200\ m, \cdots, 900\ m$. The directional tolerance used for the lag is $\frac{1}{2}$ of whatever the lag distance is (here, $100\ m$). Therefore, the first *bin* will cover the interval ($50\ m - 150\ m$) with center at $100\ m$. The second *bin* will cover the interval ($150\ m - 250\ m$) with center at $200\ m$, etc. Suppose we want to compute the semivariogram at $h = 100\ m$. Because this is an omnidirectional variogram, we can draw two circles around each point with radius $50\ m$ and $150\ m$ and pair this point with all the points that fall in the area between the two circles. For example, in the figure below, point 1 will be paired with all the points in the shaded area. Similarly you do the same for every point.

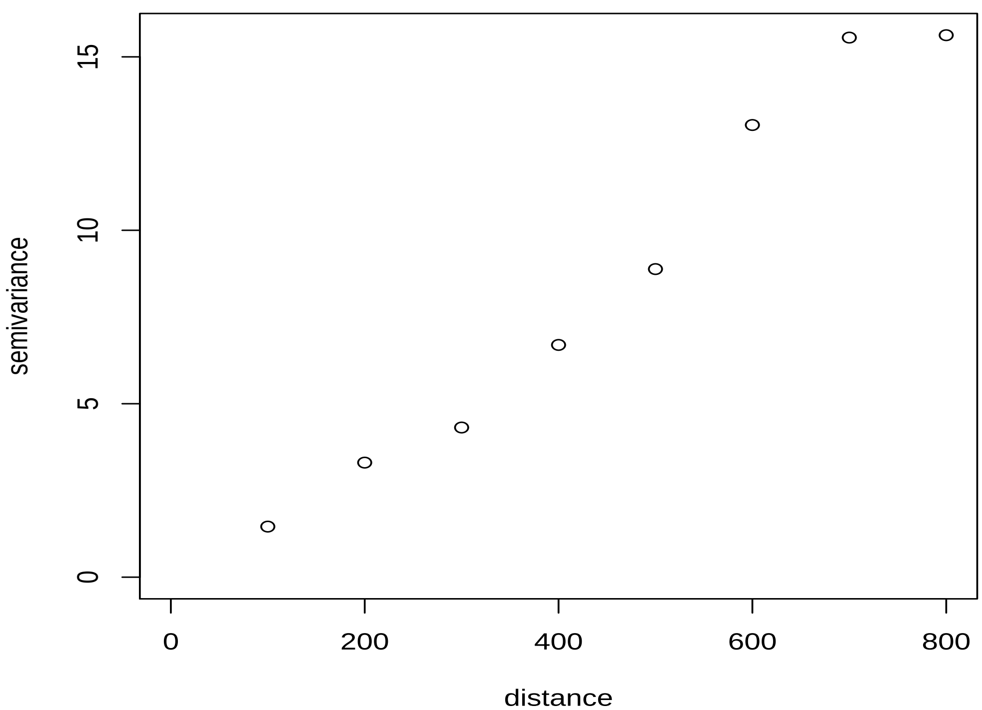


Obviously, there is no need to specify any angular tolerance when computing an omnidirectional variogram.

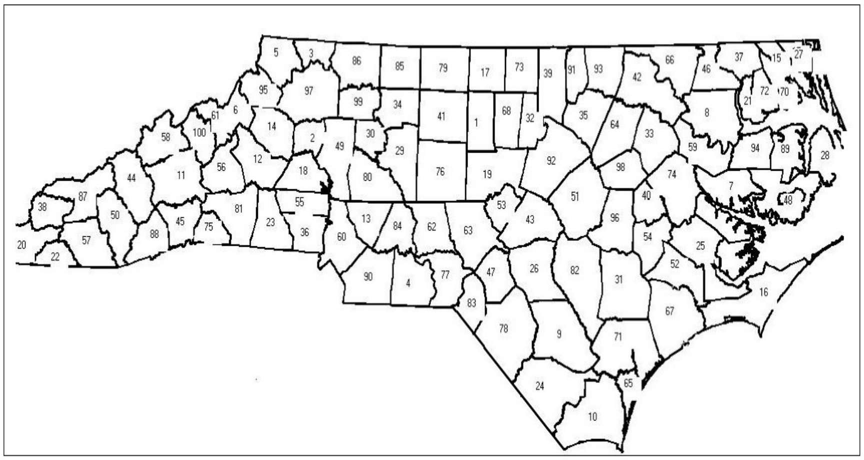
Let's compute the semivariogram on the west-east direction:

```
> var2 <- variog(b, max.dist=800, dir=pi/2, tol=0, uvec=seq(0, 800, by=100))
variog: computing variogram for direction = 90 degrees (1.571 radians)
        tolerance angle = 0 degrees (0 radians)
> var2$u
[1] 100 200 300 400 500 600 700 800
> var2$v
[1]  1.458333  3.303030  4.314815  6.695652  8.882353 13.035714
[7] 15.555556 15.625000
> var2$n
[1] 36 33 27 23 17 14  9  4
```

Here is the semivariogram plot:



North Carolina 1979-84 SIDS Data:
Sudden Infant Death Syndrome (SIDS) data are collected for each of the 100 counties of the state of North Carolina and the total number is reported for the years 1979-84. The data points are assumed to be the county seats (see figure below).



Sudden Infant Death Syndrome (SIDS), is the sudden death of any infant up to 12 months old, that is not foreseen by the family history and where the postmorten examination does not give any logical explanation as to why the infant has died. SIDS is, or at least it was until very recently, the main cause of postneonatal death. It leads to about 7000 deaths per year in the United States - it is responsible for the deaths of about two infants per 1000 live births. See, Cressie, N. (1993). *Statistics for Spatial Data*. John Wiley, New York, pp. 244-248, 386-389.

You can access the data in R as follows:

```
> a <- read.table("http://www.stat.ucla.edu/~nchristo/statistics_c173_c273/
  nc_sids.txt", header=TRUE)
```

Please do the following:

a. Plot the data using the `plot` and `points` functions.

b. You may want to transform your data.

c. Compute omnidirectional and directional variograms on each of the variables or transformed variables.

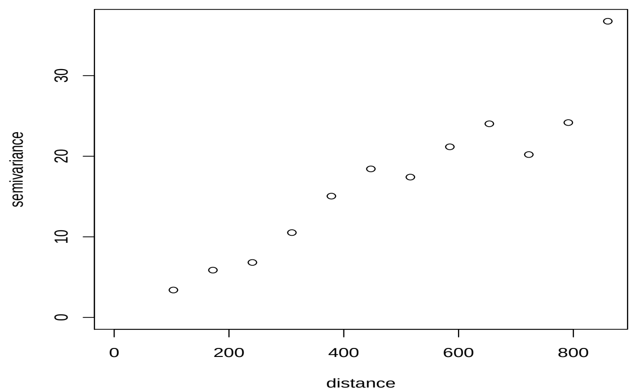
d. Compute omnidirectional and directional variograms on the sids per number of births ratio.

e. Fit (by eye) the exponential, spherical, and linear model variograms.

Fitting a variogram model over the empirical variogram:
If we know the variogram parameters we can use the function *lines.variomodel* to draw it over the empirical variogram. In practice of course we don't know the parameters. In the example here we will estimate the parameters *by eye*.

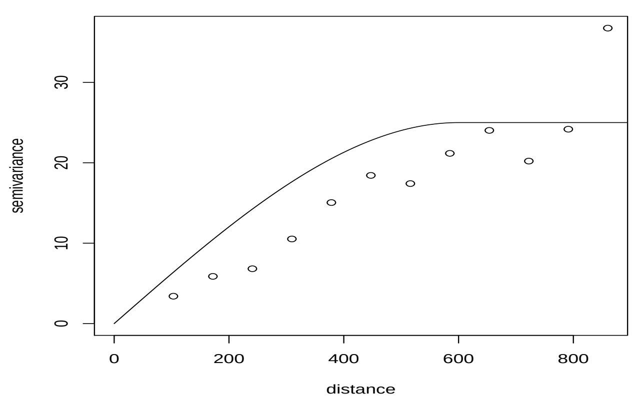
The graph below is the *omnidirectional* variogram for the data:

```
> a <- read.table("http://www.stat.ucla.edu/~nchristo/data_var.txt",
  header=TRUE)
```



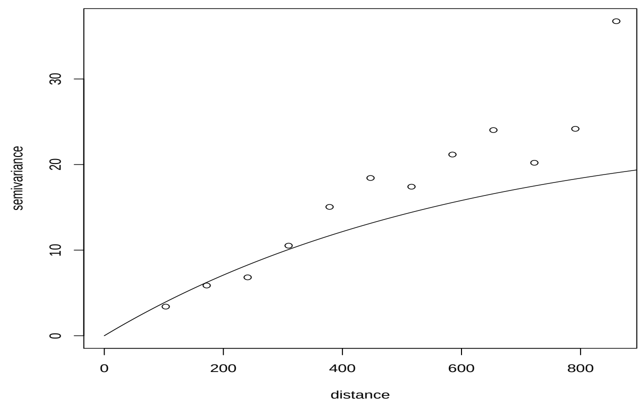
Let's assume that we want to fit the spherical variogram over this empirical variogram. We can estimate by eye that the partial sill, range, and nugget are equal to 25, 600, and 0 respectively. Here is how the *lines.variomodel* works:

```
> variogram1 <- variog(b, max.dist=900)
variog: computing omnidirectional variogram
> plot(variogram1)
> lines.variomodel(cov.model="sph", cov.pars=c(25,600), nug=0, max.dist=900)
```



For exponential varogram:

```
> lines.variomodel(cov.model="exp", cov.pars=c(25,600), nug=0, max.dist=900)
```



For the linear variogram:

```
> lines.variomodel(cov.model="power", cov.pars=c(0.035,1), nug=0, max.dist=900)
```

This can also be used for a power variogram. Instead of "1" we can use any other value between 0 and 2. The general form of the power variogram is bh^{ϕ}, where b is the slope and ϕ is the power parameter.