Lecture 9: Kriging

Dr Adams

Setting the projection

How to set the projection, when it is missing.

- Define Projection (ArcGIS)


```
spObject@proj4string <- CRS("+proj=utm +zone=17
    +ellps=GRS80 +datum=NAD83 +units=m +no_defs")
```

Projections using EPSG SRID

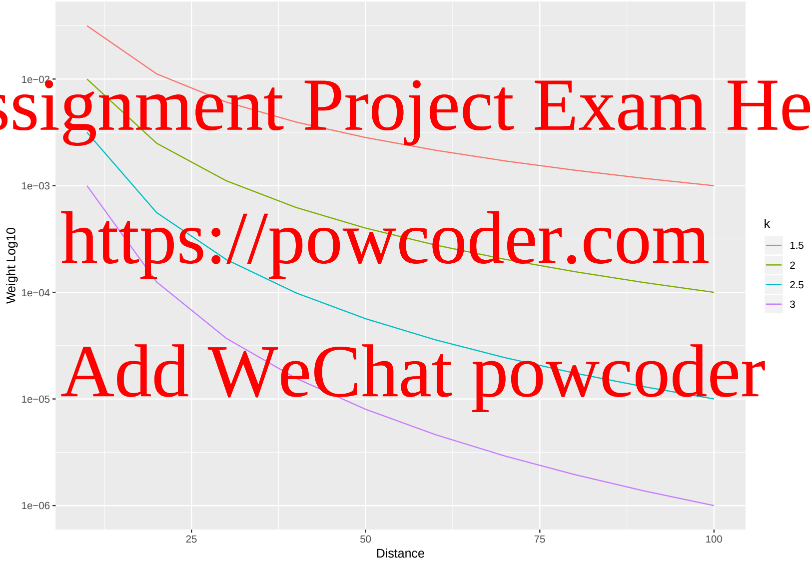
EPSG

- Spatial Reference System Identifier
- Maintained by the International Association of Oil & Gas Producers (OGP) Surveying & Positioning Committee


```
spObject@proj4string <- CRS("+init=epsg:26917")
```


http://spatialreference.org/ref/epsg/

IDW Affect of k



Deterministic Methods

Where outcomes are precisely determined through known relationships among states and events

- No random variation
- Given inputs produce the same outputs
- No uncertainty assumed

Stochastic Methods

A stochastic model will include some uncertainty in the estimate

Kriging Formula (Example)

$$Z(x) = \mu(s) + \epsilon(s)$$

Where,

$Z(x)$ is the variable of interest
$\mu(s)$ is a deterministic trend
$\epsilon(s)$ are the auto-correlated errors

Kriging

Kriging is a statistical method that uses a variogram to calculate the spatial autocorrelation between points at graduated distances

- The calculation of spatial autocorrelation is used to determine the weight of values at differing distances.
  - How much are observations related at varying distances?

Variogram

- A *function* (model) describing the degree of spatial dependence of a spatial random field

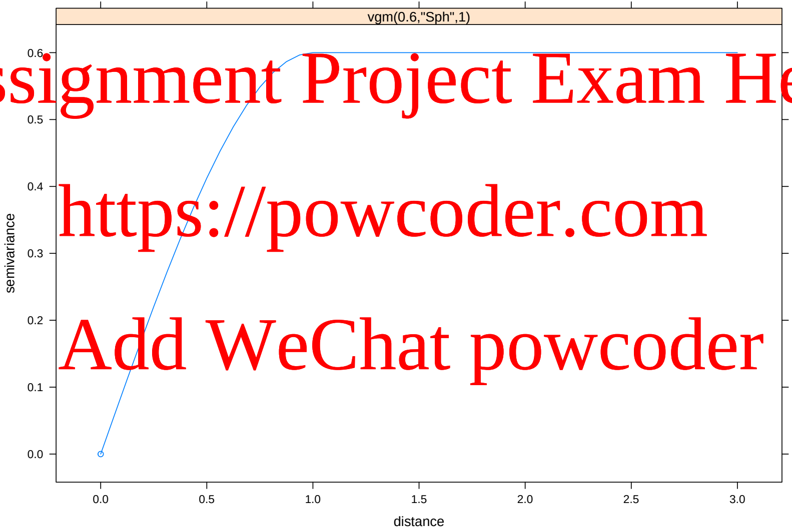
- Many functions exist for defining the relationship
- Attempts to answer:
  - How much will any two samples vary at a given distance.

Variogram II

- A visual depiction of statistical dependence between pairs of points

- x-axis: distance
- y-axis: semi-variance

Variogram Visual



Variogram III

Empirical Variogram

- Based on observed data

Model Variogram

- Theoretical/model to represent the underlying relationship.

Example Variogram Formula, Spherical Model

$$\gamma(h) = (s - n)\left(\left(\frac{3h}{2r} - \frac{h^3}{2r^3}\right)1_{(0,r)}(h) + 1_{[r,\infty)}(h)\right) + n1_{(0,\infty)}(h)$$

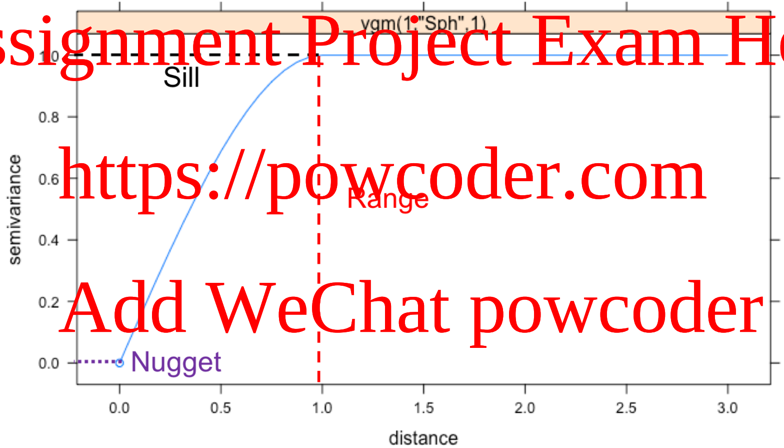
where,

s, (Sill) value at which the model flattens out
r, (Range) distance at which model first flattens out
n, (Nugget), value at which the semi-variogram (almost) intercepts the y-value.

- Expectation, distance 0 values are equal
  - Due to sampling error or randomness they may not

h, Distance

Sill, range, nugget



Sill, range, nugget

Most variogram models are defined by their sill, range and nugget.

Range

- Distance where the variogram reaches the sill
- Distances greater than the range do not exhibit spatial-autocorrelation

Sill

- Represents the variance of the random field.

Variance

- Variance: Expectation of the squared deviation of a random variable from its mean

$$Var(X) = E[(X - \mu)^2]$$

X is a random variable
E is the expectation (long run mean) μ is the mean of X

or

$$\mathrm{Var}(X) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2$$

Work out the Varaince

$$\mathrm{Var}(X) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2$$

$$X_1 = 1, 20, 22, 15, 30$$

$$X_2 = 10, 11, 13, 15, 17$$

$$X_3 = 1, 100, 220, 300, 500$$

Variogram Objective

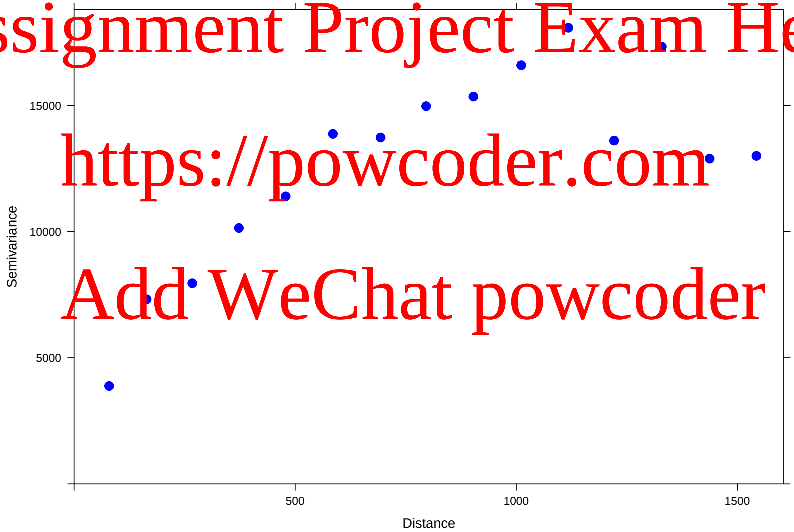
- Define covariance between points relative to distance
  - Conditional variance on distance
- The measure applied is semi-variance

Semivariance

- Prior to selecting a variogram model, it is necessary to understand the semivariance from observed (empirical) data.
- Empirical data is unlikely to have values at every distance
  - Instead values are binned
    - Think Histogram

Variogram Example

```
lead.vgm <- variogram(logLead~1, data = meuse)
```

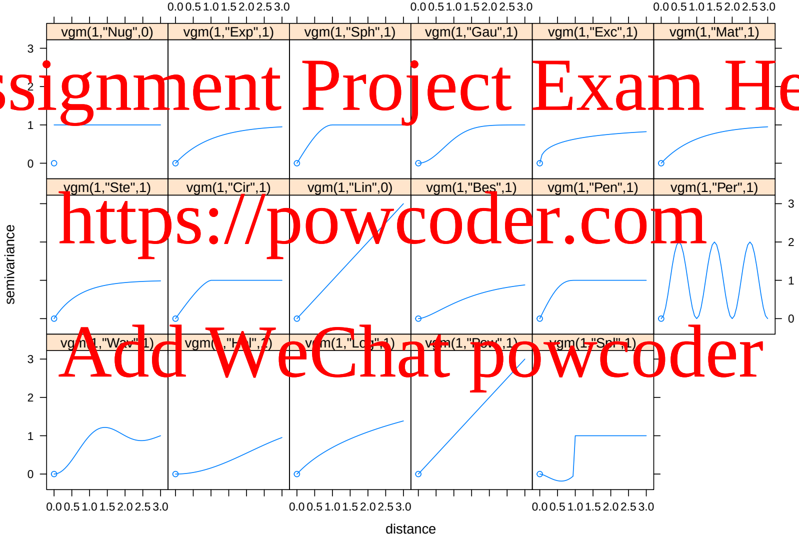
From Emipircal to Model

- Review the empirical variogram
- Select model with similar shape
- Fit model to empirical data

Objective

- Trying to select a model that best represents the relationship between distance and semivariance.

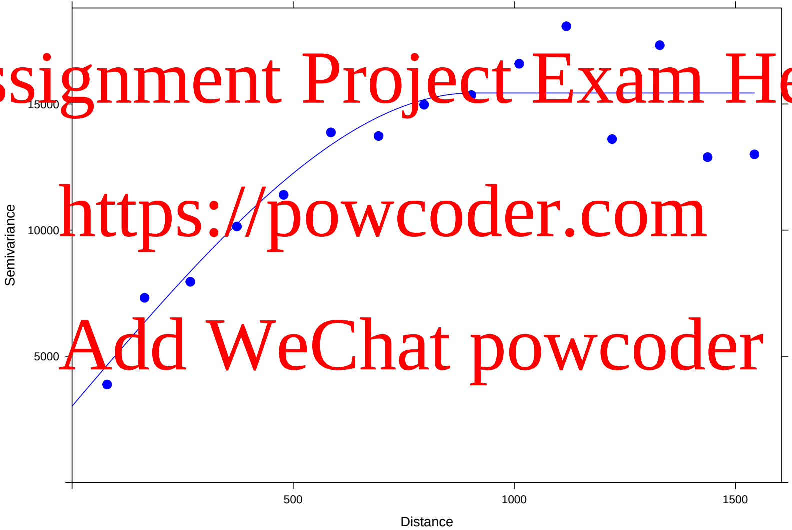
Many Choices



Fit model to data Code


```
lead.fit <- fit.variogram(lead.vgm, model = vgm("Sph"))
plot(lead.vgm, lead.fit)
```

Fit model to data



Variogram Fitting Challenges

- Approaches can use a variogram cloud
    - Very large number of data points

$\frac{n^2}{2}$

- Nonlinear models
    - Require estimated starting values

Parameters


```
lead.fit
```



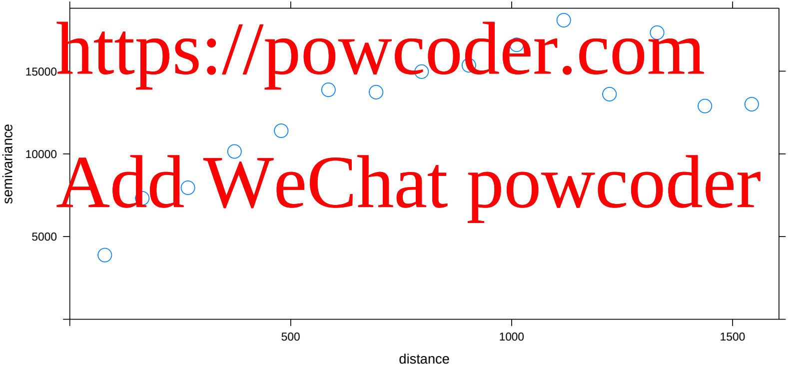
```
##   model     psill    range
## 1   Nug  3024.016   0.0000
## 2   Sph 12412.721 906.6158
```

How are the parameters estimated?

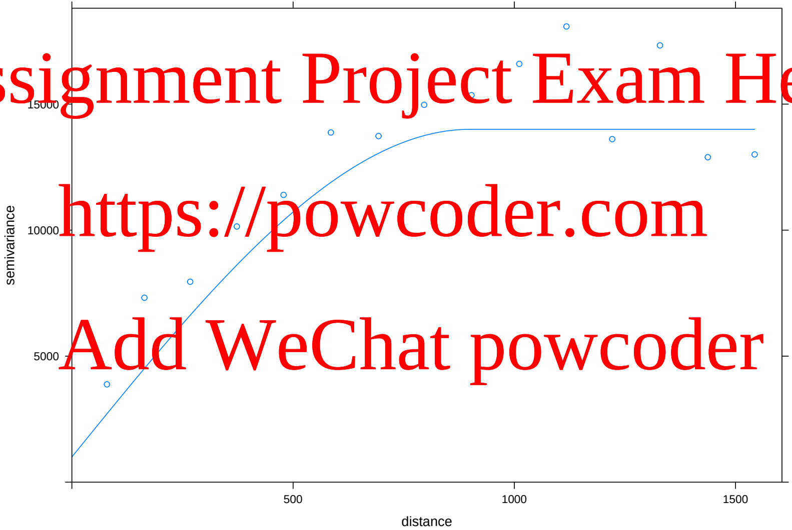
- Initial estimate
  - Nugget
  - Range
  - Sill
    - In R, Partial Sill: Sill - Nugget
- Optimization

Inital Guess

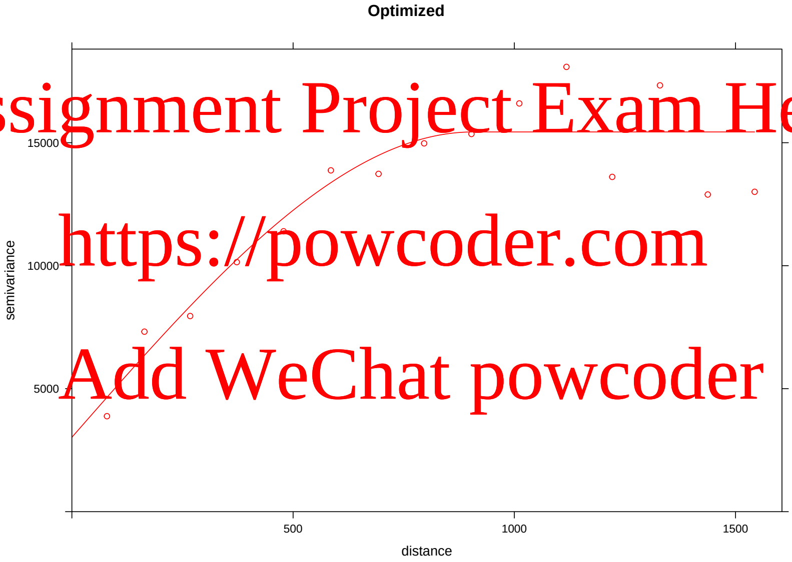
- nugget = 1000
- sill = 13000
- range = 1000



Initial Fit



Optimized


Optimized
semivariance
distance
15000
10000
5000
500
1000
1500

Optimized Parameters

Initial


```
##   model psill range
## 1   Nug  1000     0
## 2   Sph 13000   900
```


Optimized


```
##   model     psill    range
## 1   Nug  3024.016   0.0000
## 2   Sph 12412.721 906.6158
```

Minimzing Sum of Squared Errors

Initial


```
attr(lead.fit, "SSErr")
```

```
## [1] 26392.31
```


Optimized


```
attr(lead.fit.intial, "SSErr")
```

```
## NULL
```

Weights Review

- IDW
  - Weighting is determined by the exponent and distance
    - Linear change

- Kriging
  - Weights are determined by the variogram model and distance
    - Non-linear relationships possible

Kriging Details

Two step process

1. Determine the spatial covariance structure in the data.
2. Interpolate unobserved locations using weights from the variogram model.

Kriging Assumptions

1. Data are normally distributed.
2. Stationarity, spatial processes do not vary across the field.
3. Isotrophy
   - Process is uniform in all directions
     - Violation: anisotrophy

Testing Assumptions

1. Data are normally distributed
   - Histogram, apply necessary transformation
2. Stationarity
   - Spatial processes do not vary across the field.
     - Mean
     - Variance
3. Isotrophy
   - Uniform in all directions

Stationarity Evaluation

- No formal tests

Visual Approach

- Divide the area up into polygons
  - Calculate the mean and variance by polygon

Stationarity Evaluation Process

1. Extent of the observations (xmin, xmax, ymin, ymax)
2. Create a grid across the study area
3. Calculate the mean and variance within raster cells.

Stationarity Example

Get the extent

```r
# Find extent values
# Using functions in the raster package
library(raster)
area <- extent(meuse@bbox)

area
```

```
## class       : Extent
## xmin        : 178605
## xmax        : 181390
## ymin        : 329714
## ymax        : 333611
```

Create a grid

```r
# Create a raster of 5 x 5 covering the area
raster_area <- raster(nrows = 5, ncols = 5, ext = area,
                      vals = -999)
# Convert Raster to Polygons
polygons_area <- rasterToPolygons(raster_area)
```


Count, Mean & Variance

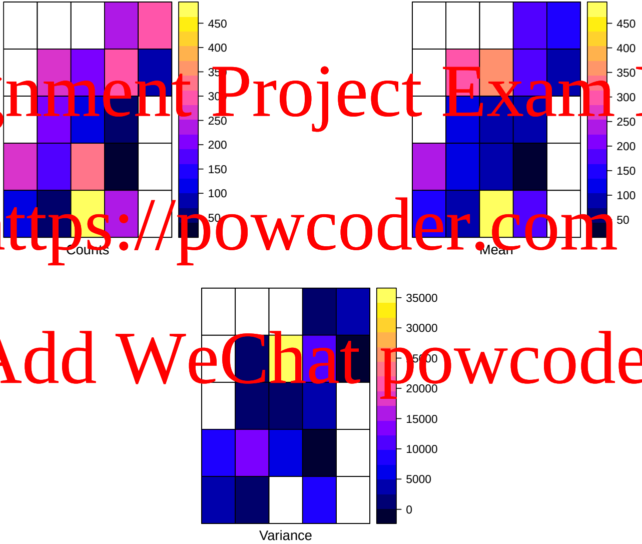
```r
# Removing all columns in meuse except logLead
meuse2 <- meuse[,-(1:12)]

# Number of obserations per polygon
assessment <- over(polygons_area, meuse2)

# Mean of values
assessment$mean <- over(polygons_area, meuse2,
                        fn = mean, na.rm = T)[,1]

# Variance of Values
assessment$var <- over(polygons_area, meuse2,
                       fn = var, na.rm = T)[,1]
```

Clean up

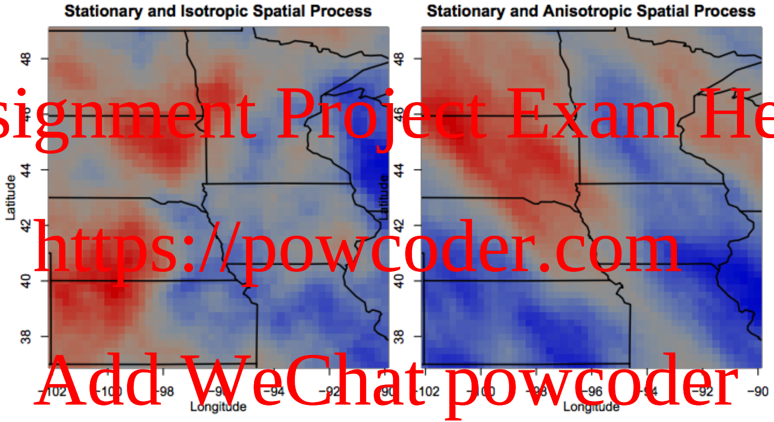
```r
# Change Names
names(assessment)[1] <- c("counts")

# Replace Polygon Data
polygons_area@data <- assessment

summary(polygons_area@data)
```

```
##      counts          mean            var
##  Min.   : 39   Min.   : 43.5   Min.   :   40.5
##  1st Qu.:112   1st Qu.: 85.0   1st Qu.: 1972.5
##  Median :210   Median :145.8   Median : 3611.7
##  Mean   :203   Mean   :170.1   Mean   : 6383.9
##  3rd Qu.:276   3rd Qu.:187.0   3rd Qu.: 8243.0
##  Max.   :464   Max.   :464.0   Max.   :34189.5
##  NA's   :8     NA's   :8       NA's   :9
```

Plots


450
400
250
200
150
100
Counts
450
400
350
300
250
200
150
100
50
Mean
35000
30000
20000
15000
10000
5000
0
Variance

Isotrophy

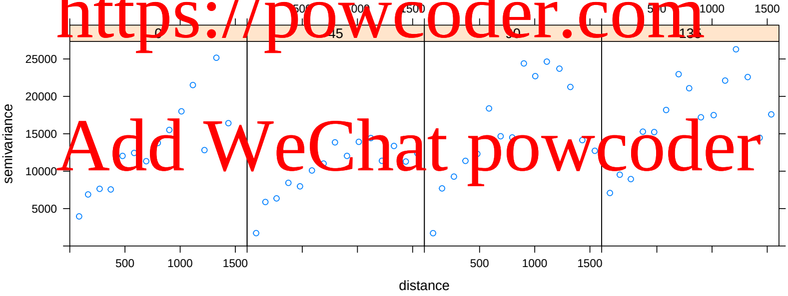


Spatial dependence is stronger in the NW-SE direction than the SW-NE direction

- library(SpTest) Vignette Example

Directional Variograms

```
lead.vgm.dir <- variogram(logLead~1, data = meuse,
                          alpha=c(0,45,90,135))
plot(lead.vgm.dir, threshold = 30)
```

Kriging Methods

- Simple Kriging
- Ordinary Kriging
- Universal Kriging
- Co-kriging
- Poisson Kriging

Simple Kriging

- All kriging assumptions
- Assumed $\mu = 0$
- Little value to empirical science

Ordinary Kriging

- Assumes stationarity of the field
  - Constant Mean and Variance

- Trends across the study area cannot occur.
- Any value of μ for the field.

Universal Kriging

- Relaxes the stationarity assumption for mean values
  - Allows the mean to vary in a deterministic way
    - Trend across the study area
- Second-order stationarity
  - Variance must be stationary

Co-kriging

- Includes an independent variable that is correlated with variable of interest.

- For example, rainfall predictions
  - Include elevation (mountainous)

Poisson Kriging

- Incidence counts or disease rates

Kriging Limitations

- Misspecification of the variogram model can lead to erroneous results.
- Challenging to evaluate assumptions
- Difficult to meet certain assumptions
- Accuracy is challenging with few points